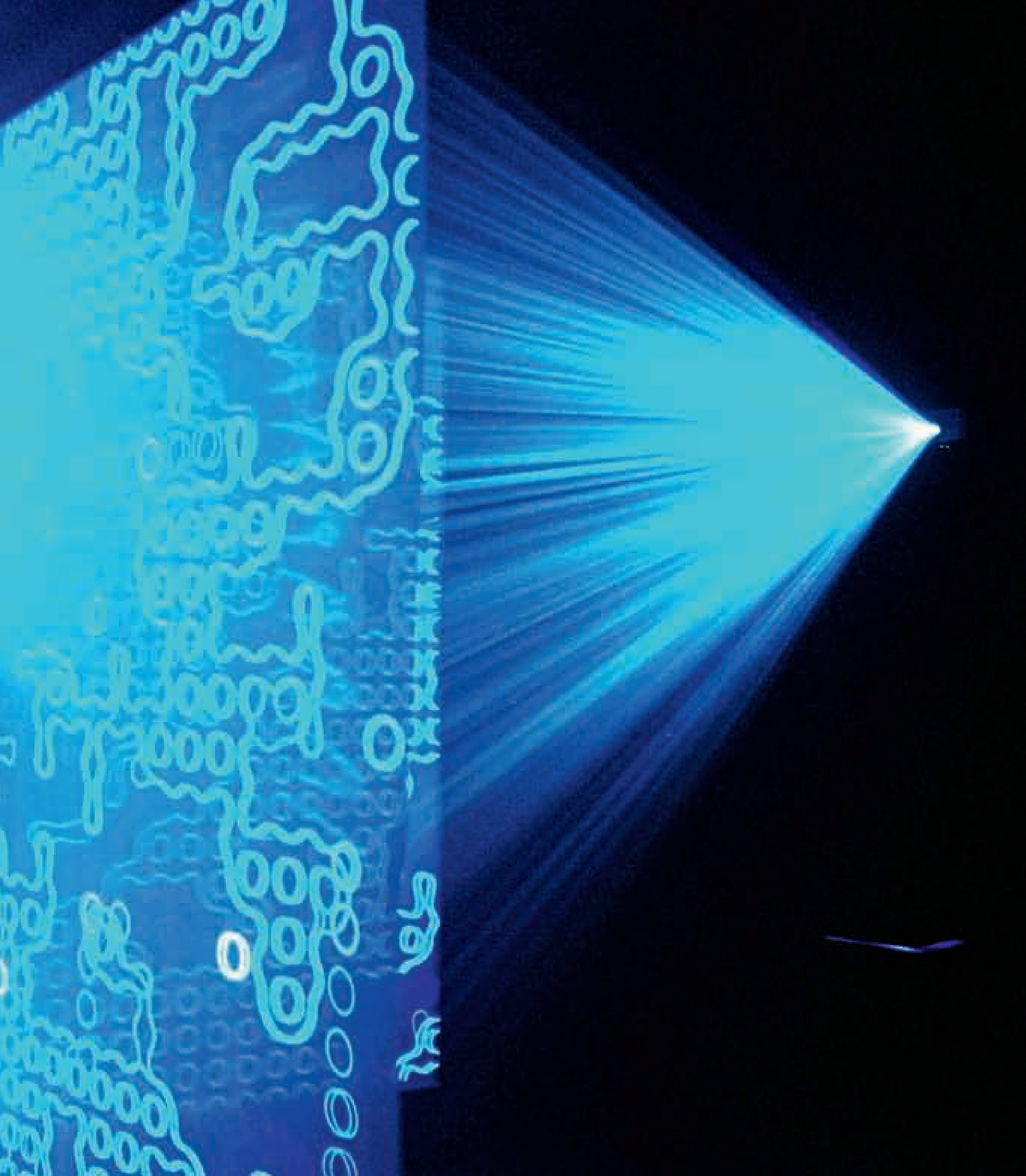
INNOVATION
Perspectives for the 21st Century

BBVA



INNOVATION
Perspectives for the 21st Century

For this third book in the BBVA series, we have chosen innovation as the central theme. It was chosen for two fundamental reasons: the first was the decisive importance of innovation as the most powerful tool for stimulating economic growth and improving human standards of living in the long term. This has been the case throughout history, but in these modern times, when science and technology are advancing at a mind-boggling speed, the possibilities for innovation are truly infinite. Moreover, the great challenges facing the human race today—inequality and poverty, education and health care, climate change and the environment—have made innovation more necessary than ever. Our economy and our society require massive doses of innovation in order to make a generalised improvement in the standards of living of nearly 7 billion people (the number continues to grow) compatible with the preservation of the natural environment for future generations. The second reason for choosing this theme is that it is consistent with BBVA's corporate culture. Our group's commitment to the creation and dissemination of knowledge ties in directly with the vision that guides every aspect of our activity: "BBVA, working towards a better future for people." People are the most important pillar of our work, and the work we do for and on behalf of people is supported by two other pillars of our culture and strategy: principles and innovation.

Index

Innovation for the 21st Century Banking Industry

Francisco González

Chairman of BBVA

This book, *Innovation: Perspectives for the 21st Century*, is the third in a series of annual publications by the BBVA Group. The motivation behind these publications is to publish expert knowledge on the key issues shaping the future course of the 21st century and relay this knowledge to society. To this end, we seek out leading researchers and creative minds from around the world and ask them to address rigorously and objectively the latest developments in knowledge and the ongoing debates on research and artistic creation in their own fields, using a language and approach that non-specialised readers can understand.

The first book in the series, *Frontiers of Knowledge*, was published in connection with the institution of the Frontiers of Knowledge Awards granted by the BBVA Foundation. It focused on recent breakthroughs and key challenges in each of the eight award categories: biomedicine; ecology and conservation biology; climate change; information and communication technologies; economics, finance and business management; development cooperation; and the contemporary arts.

The second book offered a comprehensive overview of the complex phenomenon of globalisation, which today deeply affects every aspect of people's lives.

In order to give the collection a sense of continuity, we have chosen innovation as the central theme of this third book. It was chosen for two main reasons: the first was the decisive importance of innovation as the most powerful tool for stimulating economic growth and improving human standards of living in the long term. This has been the case throughout history, but in these modern times, when science and technology are advancing at a mind-boggling speed, the possibilities for innovation are truly infinite. Moreover, the great challenges facing the human race today—inequality and poverty, education and health care, climate change and the environment—have made innovation more necessary than ever. Our economy and our society require massive doses of innovation in order to make a widespread improvement in the standards of living of nearly 7 billion people (the number continues to increase) compatible with the preservation of the natural environment for future generations.

The second reason for choosing this theme is that it is consistent with BBVA's corporate culture. Our group's commitment to the creation and dissemination of knowledge ties in directly with the vision that guides every aspect of our activity: "BBVA, working towards a better future for people". People are paramount in our work, and the work we do for people is supported by two other pillars of our culture and strategy: principles and innovation.

BBVA's principles can be summed up in the belief that ethics are not only desirable but also profitable. Acting in accordance with strong values of honesty, integrity and transparency is essential for establishing a close and lasting relationship of trust with all our stakeholders: our shareholders, suppliers, regulators and, above all, our employees and our customers.

This ethical commitment extends to all the societies in which we operate and to society as a whole because we believe that economic development and social stability are the keys to ensuring BBVA's continuous, profitable growth . For this reason, BBVA is actively involved in a variety of social projects, with a particular focus on promoting education and knowledge. This is the context that has inspired the publication of these books as well as a host of other initiatives, such as the Frontiers of Knowledge Awards and the different activities organised by the BBVA Foundation, in addition to ambitious educational programmes launched by the bank in every country in which we have a presence.

However, although these actions are undeniably important, at BBVA we believe we make our greatest contribution to improving people's lives through diligently performing our daily activity. The banking industry, and the financial industry in general, carry out tasks essential to people's everyday lives and to economic development and social stability. BBVA strives to offer its customers a wider and better range of solutions each day, and to make these solutions available to more and more people. Innovation is a vital tool; with it, our daily efforts can achieve the best results, and BBVA can become, as we want it to be, the best universal bank worldwide—in other words, the bank that offers the best and most varied solutions for people and for companies.

I will return to this later, but first I would like to say that the undertaking of publishing these books over the past three years has been an extremely gratifying experience. Each year we have been privileged to work with some of the world's leading experts on truly fascinating subjects. The opportunity to interact with them and their ideas has enriched us all.

This year, again, I am very proud of the calibre of the authors who kindly agreed to participate in our project. The book boasts texts written by a select group of the world's best and most prestigious experts in their respective fields. Some are repeating the experience—like Professor Rosenberg, who contributed a magnificent article on globalisation to last year's publication—which constituted a great honour and show of support for our project. Others have collaborated with BBVA on different innovation projects, or represent institutions with which we have signed agreements or established partnerships in this field. And all have made valuable contributions that give us a glimpse of the "state-of-the-art" in innovation.[1]

I wish to express my gratitude and that of the entire BBVA Group to all the authors for participating in this book entitled *Innovation: Perspectives for the 21st Century*. On reading it, I trust that all of you will find a rich, varied and thought-provoking discussion of innovation--a complex phenomenon of paramount importance for the society and economy of the 21st century.

[1] Here I would like to offer a special tribute to the memory of Professor Chris Freeman of the University of Sussex, one of the pre-eminent contemporary experts on economic cycles and the economics of innovation. His untimely passing this summer precluded the possibility of recruiting him for this project, in which a number of his colleagues, collaborators and disciples have participated.

AN OPEN, PLURALISTIC VIEW OF INNOVATION

Innovation is extremely hard to define. Schumpeter, the great economist who positioned innovation at the centre of the economic debate, made a distinction between invention, defined as the manifestation of a new idea or a previously unknown fact, and innovation, the ability to successfully apply that idea in practice (Schumpeter, 1934). Thus, innovation can be very loosely defined as "a change in the thought process for doing something, or the useful application of new inventions or discoveries" (McKeown, 2008).

Working on the premise of these general ideas, innovation has been analysed in countless contexts and from very different perspectives—always associated with "positive change"—in such disparate fields as technology, economics, business, sociology, the arts or the multiple branches of engineering.

Schumpeter supplied us with a definition of economic innovation which lists the different forms that innovation can take:

1. the introduction of a new good or service;
2. the introduction of a new method of production;
3. the opening of a new market;
4. the conquest of a new supply source of raw materials or semi-manufactured goods;
5. the implementation of a new organisation in any industry.

The importance of innovation in the field of economics has even given rise to a discipline known as Neo-Schumpeterian Economics, which views all economic development as the result of innovation (see, for example, Freeman, 1982).

Every innovation is the result of a process in which the idea for a possible practical application of an invention is first posited and subsequently developed until it is ready to be introduced into the market. This is why economic and business texts dealing with innovation are just as concerned with the "sources of innovation" and the processes that encourage the generation of potentially useful ideas as they are with the best mechanisms, structures or incentives for transforming those ideas into goods and services that can effectively create value on the market.

The connection between science, technology and innovation, the relationship between academic research and companies, public policy and the role of the market, government and corporate structures, business management in the area of innovation—all of these subjects have been the focus of special attention in the literature on innovation. This diversity of themes and their complexity explain both the importance and the difficulty of developing a Theory of Innovation (see Nelson and Winter, 1977).

A wide range of scholarly approaches to the concept of innovation is documented in Fagerberg (2004), which I recommend to interested readers as a helpful reference for becoming familiar with the prolific and varied bibliography on innovation.

Given the diversity of approaches, aspects and ramifications of innovation and its practical applications, in this book we have chosen to offer a very wide spectrum of articles addressing the most relevant aspects of innovation, all written by authors at the very top of their respective fields.

The first segment of essays, which provide crucial insights for understanding innovation, focuses on its deepest roots. Sandy Pentland traces the roots of creativity—and, by extension, innovation—back to their source in biology, and shows how communication and interaction among the members of each species, including (of course) humans, are essential for development. Sander van der

Leeuw points out that the cognitive abilities of the human brain do not seem to have changed in the last 50,000 years; however, thanks to cultural elements (in other words, the experience of learning to exploit those abilities to the fullest), combined with advances in information and communications, the human species still has plenty of room for improvement in terms of managing its natural environment, and that improvement is made possible by innovation.

The second segment of essays focuses on the institutional aspects of innovation. Nathan Rosenberg discusses the complex relationship between science and technology. The traditional view is that science "leads" and technology "follows", but Professor Rosenberg points out that technology is much more capable of "explaining" scientific progress than we have been led to believe. Hiroyuki Itami underscores the role that organisations (corporate, government, or non-profit) play in knowledge accumulation mechanisms, which are vital for innovation, while Alfonso Gambardella focuses on how market mechanisms can encourage innovation—mechanisms that are primarily fuelled by the utilisation of that accumulated knowledge.

Francisco Louça, on the other hand, demonstrates the importance of the intricate network of cultural values, social interactions and institutions of each society for understanding innovation processes, and discusses how the match or mismatch between socio-institutional systems and the degree of techno-economic development in each period determine the long or Kondratiev waves of economic growth and recession. Based on this reasoning, the current crisis can be chalked up to the inability of economic structures, institutions, regulations and social values to keep pace with the technological revolution we are experiencing.

David Mowery's essay offers an analysis of the US National Innovation System over more than a century and concludes that its results have largely depended on the decisions made by private companies. For this reason, public innovation policies should remain consistent over time and seek the approval and support of the private sector.

Using data compiled by European economists, Edward Lorenz and Bengt-Åke Lundvall empirically prove that the structural traits of economies, like education systems and labour market structures, have a significant impact on creativity, and therefore on innovation.

The third section of essays examines innovation from a "micro" perspective, exploring how innovations are generated and what plans and mechanisms should be introduced in organisations to generate and disseminate ideas and, above all, to turn those ideas into innovative goods and services on the market.

Alice Lam analyses the organisational aspects of the innovation process and points out the need to cultivate the learning and knowledge-building capacity of human resources, but also to design flexible organisations that can adapt to new technologies and processes.

In a revision of the conventional "producer-centred" approach to innovation analysis, Eric von Hippel reveals how users have become an important source of innovation thanks to advances in computer science and improved connectivity. Consequently, it is essential for companies to maintain a constant dialogue with users and devise mechanisms for working with them and making the most of their abilities.

Frank Moss describes the key cultural aspects of the MIT Media Lab's "research style", one of which is the creation of an environment of creative freedom that encourages people to ask bold questions, where failure is perfectly acceptable and where learning is an integral part of the creative process (learning by doing).

The essay written by Curtis Carlson illustrates the "best practices" for innovation developed at the Stanford Research Institute, designed to improve the odds of success for the innovative efforts of organisations.

Harry West focuses on "radical innovation" and outlines the principal elements of Continuum's process for designing and developing this type of innovation, which is the hardest to standardise but has a much greater impact.

Pascal Soboll offers a complementary perspective—that of the consulting firm Ideo—on how to create an innovative culture in organisations, in such a way that innovation is a priority for all departments rather than the exclusive concern of a small group of "experts".

This third section concludes with Joaquim Vilá, who highlights the pivotal role that senior executives must play in implementing the changes required for an organisation to achieve a robust innovative culture, and enumerates the fundamental cultural factors that these executives must embrace and preach by example.

The fourth and final segment is dedicated to the application and impact—present and/or future—of innovation in a number of relevant areas, sectors or activities.

Manuel Mira Godinho shows us how innovation is largely responsible for the reduction of extreme poverty in the world over the past several decades, and how it can continue to eradicate this problem in the future. For this to happen, developing nations must acquire policies and tools that will allow them to obtain know-how, compete for R&D funding, and join the global institutional framework, including the necessary policies for handling environmental problems.

As one might expect of a book published by a financial institution, this volume pays special attention to innovation in the service industry, and more specifically in the world of finance. Ian Miles points out that conventional academic publications on innovation have tended to focus on manufacturing industries; yet services represent a very large (and constantly growing) chunk of the economy, and innovation in this sector presents distinctive traits and demands. Developing cross-disciplinary teams is a necessity in the service industry, because innovations in this field usually involve the combination of multiple goods and services, requiring knowledge of technologies, institutions, regulations and social habits as well as of specific types of customers and customer interfaces.

Robert Litan reviews the history of financial innovation in recent decades and concludes that in many cases it has had a positive effect, similar to that achieved by innovation in any other industry, generating goods and services that are better, cheaper and delivered to the buyers more quickly. It is only when financial innovation focuses on the search for mechanisms to increase leveraging to dangerous levels that it has a negative impact, as the recent crisis clearly proves. Therefore, the competent authorities must introduce policies that will prevent the proliferation of "destructive" financial innovations without stifling true, positive financial innovation.

Xavier Vives picks up where Litan leaves off, discussing the role that financial innovation has played in the crisis and reminding us that every major technological change (the railway or the automobile in their day and the internet in ours) has been accompanied by a speculative bubble. Nevertheless, innovation—

particularly financial innovation—is essential for economic development. Consequently, we must concentrate our efforts on designing appropriate regulations for the development of financial innovations that will bring private incentives into line with general welfare interests.

Edward Rubin views innovation as a fundamental tool for solving the problem of climate change. If we hope to achieve the international goal of stabilising the levels of greenhouse gases in the atmosphere, we will have to apply technologies that are still being developed or have not even been invented on a massive scale. Public policy should focus on providing the proper incentives.

Takanori Shibata addresses innovation in the medical field, discussing the development of robots that can be used as pets and have proven to be therapeutic. Hugh Herr and Ernesto Martínez-Villalpando explain the tremendous potential of technological innovations for improving the quality of life of the 650 million people around the world who live with some kind of physical or mental disability.

Carlo Ratti and Nashid Nabian show how innovation can generate "intelligent" cities that will provide access to useful information in real time and a platform for collaboration among their inhabitants, radically improving their quality as living and working spaces.

Finally, Tod Machover explores the applications of innovation in music through musical "tools" that allow anyone to make music. This technology has great therapeutic potential; however, on a more general level, it also offers a new model of interaction between people and music that is much more direct and creative.

INNOVATION IN THE FINANCIAL INDUSTRY

The financial industry is already caught up in an intense, inevitable process of transformation, and the driving forces behind it—technological progress and the social changes it is bringing about—are equally intense and inevitable.

We are currently witnessing the most disruptive technological revolution since the advent of the Industrial Revolution two centuries ago. The difference is that only a small portion of the world took part in the technological progress of the 18th and 19th centuries, while today's revolution is spreading like wildfire across the entire planet. The reason for this is simple: ours is not a revolution of the tangible (production or transport of goods) but of the intangible. It is a revolution of information. The cost of collecting, storing, processing and sending information is falling rapidly. And just as important—or perhaps even more so—is the fact that these new possibilities are within reach of almost everyone on the planet, thanks to the advent of personal computers, the internet and, increasingly, mobile phones.

This phenomenon is changing people's habits and behaviour in every area of their lives: the workplace, recreational activities, communication and even interpersonal relationships. Although all companies must deal with these changes in their customers' lifestyles and in the production and distribution processes, nowhere are their effects more drastic than in the service industry, where the information component carries much more weight (see Miles, 2000).

Banks are at the epicentre of this change. Technological evolution and social changes have a deeper and more direct effect on the financial industry than on most other sectors, for its basic raw materials are information and money. And money, in turn, can dematerialise and transform into accounting entries—in other words, into data that can be stored, processed and transmitted in real time and at costs so low that they are on the verge of disappearing altogether.

It is true that banking has not experienced—up until now—a transformation on a par with that undergone by other information-based sectors, such as the music industry. This is largely owing to the fact that banking has historically been a highly-regulated industry, subject to close scrutiny and control by public authorities. It is also partly due to the exceptionally benign economic and monetary climate of the past several decades, which fuelled the intense growth of financial activity and permitted a relatively high level of inefficiency in the industry, including the survival of a staggering number of financial institutions around the world—or, to put it another way, a severe excess plant capacity.

However, not only is the transformation of the industry inevitable, but it is also picking up speed with each passing day. The primary reason is that the technological revolution is introducing daily new and different ways of doing things, and increasing the potential for cutting costs, while the number of users who resort to non-traditional banking methods continues to grow.

The second reason is that the current crisis is imposing changes in various directions. Banks are perceived as the "culprits" of the recession, and with good reason, for a large number of institutions made some very serious mistakes and chose to ignore the basic principles of banking: prudence, transparency and even integrity (for a more detailed discussion of this issue, see the essays by Edward Litan and Xavier Vives). As a result of these mistakes, many banks have experienced serious difficulties which caused some to go under and others to go through a complete restructuring, generally funded by government bailouts. The colossal amount of taxpayer resources poured into to saving banks has severely tarnished the reputation of financial institutions and the entire industry in the eyes of ordinary citizens. Trust is what gives banks their competitive edge, but over the past several years they have lost much of their customers' trust, and the trust of society in general.

In addition, the crisis has triggered a process of sweeping changes in banking regulations: borrowing limits, higher capital and reserve requirements, the need for major investments to improve risk and compliance systems, etc. All of this boils down to less revenue and more expenditure—in other words, a reduction in the current and future profitability of financial institutions.

In short, banks must respond to the new demands of their customers and of society, and they must face this challenge with a damaged reputation, lower profits and slow growth in the traditional banking business. This situation calls for a radical transformation: banks must dramatically revise the way they interact with their customers and take a qualitative (not quantitative) leap forward in efficiency.

To a degree, these advances in efficiency will be achieved by a drastic consolidation of the banking sector, which has already begun. But the industry's true transformation will be effected with the widespread and, above all, intelligent use of technology as part of a sustained process of innovation.

In recent decades, banks have been among the most important users of information and communication technologies, which they have adopted with two primary goals in mind: to cut costs and streamline processes to increase profit margins, and to develop channels of communication other than the conventional branch office.

Yet the original technological platforms used by banks were first introduced several decades ago (in the 1960s and 1970s) and, in most cases, subsequent improvements in functions were developed based on different, more modern technologies, architectures and programs that were later added and/or hooked up to the old system ad hoc.

If it were possible to visualise the complete systems network of an average bank, it would probably resemble nothing so much as a plate of spaghetti: a tangled web of connections linking very different systems that have undergone a long string of changes and partial updates over time.

This situation generates high maintenance costs (for example, it is estimated that banks in the United States devote 80% of their total investment in systems to maintenance and only 20% to new developments). And, most importantly, it quickly becomes untenable given the pace at which new technologies appear and customers' habits and demands change.

Meanwhile, the internet revolution continues to spread (internet users now account for nearly 30% of the world's population). And the uses, capabilities and functions of the internet are proliferating day by day. The internet has become the leading source of information, an indispensable pastime (today Europeans spend more time online than they do watching television), and even a forum for personal relations: over 500 million people around the world now use social networks like Facebook which did not even exist until a few years ago.

With each passing day, the internet is gaining importance as a commercial and advertising space and as a place where people on opposite sides of the globe can work together as a team. The web is also the driving force behind the fragmentation of production chains which facilitates the outsourcing of services. In this field, services offered via cloud computing represent a major breakthrough in terms of universal access to data storage and processing at very low cost, and will undoubtedly have far-reaching implications.

Internet usage has also received a tremendous boost from advances in mobile phone technology. Thanks to these new devices, nearly 4.5 billion people (almost three-quarters of the human race) are "online" and have almost ubiquitous access to some level of information services, which has a tremendous effect—yet to be quantified—on productivity.

Mobile phones come equipped with increasingly more powerful and varied functions, functions that will gradually be incorporated into other devices that people can use anywhere, anytime (what has been dubbed the "Internet of Things").

All this opens up countless windows of opportunity, not only to cut costs but also, and most importantly, to increase revenue.

In the most technologically advanced countries, the challenge is to offer customers a wider array of information-based products and services—and not just of the financial variety—with a cost so nominal it is close to zero, and to do it in the way that is most efficient, rapid and convenient for users.

Technology also offers unprecedented possibilities for tailoring services to meet the users' needs and demands. To this end, the bank must provide customers with tools that will allow them to participate in the actual process of designing the service they wish to receive.

In developing nations, we find an array of truly historic opportunities: firstly, because the majority of global growth will be concentrated in these countries in the coming decades; and secondly, because only 900 million people in the world are currently bank customers, and there are over 2 billion people—most of whom live in the world's least developed countries—who do not have access to financial services. This situation exists because, in the conventional model of production and distribution, providing financial services that involve small amounts to a scattered population is not a profitable activity.

However, technology facilitates the introduction of much more efficient models for producing and distributing financial services—for example, through the use of mobile phones. In addition to opening up a huge new market for banks, such a measure would have a tremendously positive effect on the economic development of these countries and facilitate the inclusion of the most disadvantaged collectives.

The technology needed to do all this already exists and is improving every day. A new scenario of competition in the financial industry is taking shape, a stage on which new competitors will soon emerge: companies, many of them internet-based, with high brand awareness and none of the legacies encumbering banks (obsolete systems, costly physical distribution networks, etc.), and with the potential ability to introduce highly efficient models for offering financial services.

The banks that want to compete in this new league will have to undergo a profound transformation, but they do have a few competitive advantages for initiating this change, the most important being the vast amount of information they already possess about their customers. This knowledge must become the foundation for building a new business model, one that is firmly entrenched in technology.

The fundamental tool of this new model will be a much more modern and flexible technological platform capable of absorbing all that information about customers and exploring all possible points of contact with them. In this new model, the existing network of branch offices must be given a complete overhaul: physical distribution networks are only logical if they can offer the users added value and are perfectly integrated in a physical-virtual platform. This platform should allow customers/users to interact with the bank at any time, through any channel, with no interruptions or time lapses, in order to quickly obtain financial or non-financial solutions at minimal cost that are perfectly suited to their specific needs, and which customers can even help to design if they so wish.

Parallel to this technological revolution, banks must also undergo a sweeping organisational and cultural transformation—a transformation that will allow them to restore their reputation by offering transparency in dealings with customers, speed and flexibility in responding to their demands, and the creation of an innovative culture that allows them to find solutions to the new challenges which technology and social changes will continue to pose.

Only those banks which are capable of undertaking this transformation will be able to participate in the financial industry of the 21st century: an industry that will be much more competitive than in the past, but which will also present tremendous opportunities given the possibility of meeting people's needs much more efficiently and providing universal access to financial services in the world's least developed regions.

BBVA: AN INNOVATIVE PROJECT

At BBVA, long before the current crisis, we have always tried to stay one step ahead of the pack, and we have already started to build this new business model. Our project is upheld by three pillars: principles, people and innovation.

Principles are the cornerstone of our project. At BBVA, our efforts have always been guided by the premise that ethics are not only desirable but also profitable.

Sustaining a corporate culture of prudence, transparency and integrity at any cost is a difficult, time-consuming task, and in some cases it even means sacrificing short-term profits. But in the medium and long term, it is the only way to ensure a project's sustainability.

Thanks to this culture of principles, BBVA has managed to avoid the pitfalls that many of our competitors have stumbled into in the recent past, and consequently our relative position in the global banking industry has been strengthened.

The number-one priority of our project is people, just as our vision states: "BBVA, working towards a better future for people." We strive to build stable, long-term relationships with our customers, relationships of trust based on strict ethical conduct and an effort to provide them with the best solutions to meet their needs efficiently, conveniently, and at the best price.

And here is where the third pillar of our project comes in: innovation. Creating a truly groundbreaking, "customer-centric", rapid, simple and efficient model of interaction, in which the customer receives the best his bank has to offer, requires constant efforts to innovate in both the organisational and cultural arena and in the technological field. I would like to mention just a few of the initiatives we are working on at present.

If we want to offer customised solutions, the first order of business is to know our customers well. At BBVA, like all banks, we have compiled a huge amount of information about our customers. But turning that data into knowledge that can be used to design products that will meet each customer's unique needs and determine fair prices in accordance with his/her situation means that we must equip ourselves with cutting-edge technology. At BBVA we are leading the way in the application of data-mining and creating intelligent algorithms that will allow us to anticipate the future demands of our customers at any given time.

At BBVA, we have initiated a profound transformation of our distribution network. We are already the world's most efficient bank, but we continue to work on new branch office configurations that are even more efficient, streamlined and able to provide better service. Another project underway is the design and construction of the best remote channels, equipped with the best and most varied functions, so that customers can interact with BBVA in whatever way is most convenient for them and help us perfect the exact kind of service they prefer. The phone and the computer were followed by the mobile phone, and these will soon be joined by the iPad, television and any other devices to which customers have access.

At BBVA, we are moving towards a distribution model that goes beyond the contemporary concept of multi-channel communication, where the physical office is the heart of the system and the other channels are just useful accessories. We are developing a seamless physical/virtual space where customers can come and go between the branch office and the virtual world as they please and in perfect continuity.

This space will give rise to a new definition of a bank: a company that will offer other non-financial, information-based services which incorporate the users' own contributions and tap the potential of social networks. And all this will be achieved by taking advantage of the growing ubiquity and functionality of mobile phones and the ability of cloud computing to offer cheap universal access to all kinds of information-related services.

In pursuing this goal, BBVA has an important competitive advantage: a cutting-edge technological platform, a platform that goes far beyond the conventional model of banking systems and is capable of integrating all channels and all sources of information. Thanks to this platform, a customer who accesses the bank by any channel will always find the same BBVA, with the same capacities, and will be able to jump from one channel to another as he/she pleases without a hitch. This platform, which we have been building since 2007, is currently 80% complete and will be fully operational in less than two years.

BBVA has a vision for the industry's future and has been working for years to make it a reality. But it also has a strategy that combines this vision of the future with the current reality and the prospects of each market and each type of customer.

At BBVA we want to leverage the potential of our model in high-growth markets. For this reason, in addition to our strength in Latin America, we are building a solid franchise in the United States (the world's largest market), we have a strong presence in China and other Asian countries, and we are in the process of acquiring significant interests in Turkey. In this way, BBVA combines its strength in developed markets with a growing presence in emerging economies, where most of the global economic growth will be concentrated in the coming decades and where a high percentage of the population still does not have access to financial services—which means that the potential for growth in the financial sector is staggering.

Our highly efficient model, firmly rooted in technology, gives us a significant competitive edge over other banks when it comes to meeting the needs of customers in developed nations (highly sophisticated and with intensive technology usage). But it is also essential for developing simple, inexpensive models to provide large sectors of the population with access to banking (as we are already doing in Latin America with mobile phones, agents and bank cards) and to operate in huge markets where we do not have a strong physical presence.

In summary, we have already made tangible progress in our transformation process. However, we have also made other "intangible" advances—or, if not strictly intangible, at least difficult to quantify—derived from what we have learned after all these years of constant work, and these are no less important for the future of our institution.

Firstly, over the course of these years we have refined and perfected our model of innovation. The journey began back in 2004 with the creation of a Corporate Innovation Department, and our initial approach was, in relative terms, more focused on technological possibilities than on the demands of the market and/or the customers.

This centralised department laid the groundwork for evolving towards an innovation "spread" among the different areas of the group. At the same time, the people who are in direct contact with customers have become our principal source of ideas, and technology is now viewed as a tool—an indispensable one—for materialising those ideas.

Meanwhile, at BBVA we have evolved towards a more open model of innovation in which we cooperate with a variety of institutions; in fact, representatives of many of them (MIT, SRI, Continuum, Ideo, etc.) contributed essays to this book. This model also factors in the increasingly important element of customer input, opening up a new space for innovation promoted by the users themselves—which, as Von Hippel points out, is quickly becoming a major source of innovations (Von Hippel, 2005).

However, the most important thing may be that this process of "learning by doing", the practice of innovation, has brought about a profound cultural change in the people who work for our organisation. Today BBVA has more and better leaders, leaders who are spearheading the transformation of BBVA. And the entire BBVA organisation has embraced a new culture that is open, has a positive attitude towards change, and accepts and encourages flexibility, initiative, accountability, learning and knowledge as values that will give us a decisive competitive edge. This culture also responds to the growing demands and high standards of society with solid ethical principles, transparency and good governance as the keys to earning and maintaining the trust of our customers.

In short, at BBVA we have gone from a traditional corporate culture, with elements inherited from a time when banking was a semi-official, micromanaged industry, to a culture that will allow us to achieve our ambition, which is nothing less than to lead the transformation of the financial industry in the 21st century.

The end goal of this transformation is a new financial system, capable of stimulating growth and sustainable development and of offering more useful, high-quality solutions to meet the needs of more people around the world.

The force that fuels this transformation can only be the thirst for knowledge. This is the driving force behind the project of the BBVA Group, and the publication of this book was inspired by our desire to express and share that motivation.

BIBLIOGRAPHY

Fagerberg, J. (2004), "Innovation: a Guide to the Literature", in J. Fagerberg, D. C. Mowery and R. R. Nelson, *The Oxford Handbook of Innovation*, London: Oxford University Press, pp. 1-26.

Freeman, C. (1982), *The Economics of Industrial Innovation*, London: Frances Pinter.

McKeown, M. (2008), *The Truth About Innovation*. London: Prentice Hall.

Miles, I. (2000), "Services Innovation: Coming of Age in the Knowledge-Based Economy", *International Journal of Innovation Management* 14(4), pp. 371-389.

Nelson, R., and S. Winter (1977), "In Search of a Useful Theory of Innovation", *Research Policy* 6(1), pp. 36-76.

Schumpeter, J. (1934), *The Theory of Economic Development*. Boston, MA: Harvard University Press.

Von Hippel, E. (2005), *Democratizing Innovation,* Cambridge, MA: The MIT Press

BBVA

The Roots of Innovation

Alex Pentland
Massachusetts Institute of Technology (MIT)

Decades of research in social psychology have captured the surprising ability of people to "read" one another. In contexts as different as evaluating classroom teachers, selecting job applicants or concluding jury deliberations, human judgments are made on the basis of extremely thin slices of observational data. Across a wide range of studies, psychologists find that research subjects on average accurately predict outcomes in such pursuits 70% of the time. That success rate holds when predicting end results occurring days, weeks or even months later.

How is this possible? My theory is that our ability to read each other starts with what are known in biology as *honest signals*. Evolutionary models predict that all social species are likely to develop honest signals, a reliable communication system that develops to coordinate behavior between individuals. Typically the signals include gestures, expressions or calls. Not only are they usually trustworthy cues, honest signals are also unusual because they appear to trigger changes in people receiving signals that are advantageous to the people who send them.

It's likely that human ancestors used such signals to coordinate their actions long before sophisticated human language evolved. A relative newcomer in hominid evolution, language was likely layered upon older primate signaling mechanisms that used social network strategies to find resources, make decisions and coordinate group action. By better understanding their influence today, we can shed light on the structure and function of modern social networks. For instance, honest signals can increase the energy level within a hunting team or, for that matter, a creative team through contagious excitement. They create a more cohesive family group by increasing empathy and trust through mimicry signaling.

When we watch a conversation between two people and carefully measure the timing, energy and variability of the interaction, we find several examples of honest signals. My research group concentrates on four components of this human signaling. *Mimicry* is the reflexive copying of one person by another during a conversation, resulting in an unconscious back-and-forth trading of smiles, interjections and head nodding. *Activity* indicates interest and excitement, familiar to us from the connection between excitement and the activity level of children. *Influence* of one person over another can be measured by the extent to which one person causes the other person's pattern of speaking to match theirs. And *consistency*, or fluidity, of speech and movement is perceived as a marker of expertise.

To measure the impact of these very ancient social signals, we developed some very modern tools and practice what we call *reality mining*. We collect data mostly with custom-designed electronic badges and sometimes with "smart" phones and other electronic devices. The instruments uncover and quantify the role that some social signaling mechanisms play in everyday decision-making. By examining the back and forth of signaling behavior in dyads and small groups—paying no attention to words or the identity of individuals—we can accurately predict outcomes of speed-dating encounters, job interviews, even salary negotiation outcomes to within $1,000. In a wide variety of situations ranging from business management to first dates to the effects of political opinion, we find that roughly 40% of variation in outcomes can be attributed to signaling-based models of social information processing. That is equivalent to the estimated influence of genetic makeup on individual behavior and is far too large, we believe, to ignore.

INFLUENTIAL COMMUNICATION

Honest signals influence critical activities including negotiation, group decision making and group management. In fact, they are accurate predictors of human behavior. For example, if one member of a group is happy and bubbly, others will tend to become more positive and excited—an effect known as mood contagion. Moreover, this signaling-induced effect on mood serves to lower perceptions of risk within groups and increase bonding. Similarly, people tend to mimic each other automatically and unconsciously. Despite being unconscious, this mimicking behavior has an important effect on participants: It increases how much they empathize with and trust each other. Not surprisingly, negotiations with lots of mimicry tend to be more successful, no matter which party starts copying the other's gestures first.

Each of these signals likely has roots in biology, specifically in our brains. Mimicry is believed to be related to cortical mirror neurons, a distributed brain structure that seems to be unique to primates and is especially prominent in humans. Mirror neurons react to other people's actions and provide a direct feedback channel between people. Newborns, for instance, mimic their parents' facial movements despite their general lack of coordination. Similarly, our activity level is related to the state of our autonomic nervous system, an extremely old neural structure. Whenever we need to react more vigorously—say in fight-or-flight situations or when sexually aroused—this system increases our animation levels. On the other hand, we tend to be listless and less reactive when our autonomic nervous system is blunted, as during clinical depression. The relationship between autonomic nervous system function and activity level is sufficiently close to enable us to use it accurately to estimate the severity of depression.

THE HABITUAL AND ATTENTIVE MIND

How do social signals interact with language? Evolution rarely discards successful working parts. It generally either builds additional structures while

> “The ability to go beyond association-based learning may be the key contribution that the attentive mind makes to the fitness of our species”

retaining the old capabilities or subsumes old structures as elements of the new. When our language capabilities began to evolve, our existing signaling mechanisms most likely were incorporated into the new design. The question, then, is how has modern human society been shaped by our ancient signaling mechanisms, and to what extent do these mechanisms still govern our lives?

A partial answer to this question can be found in the work of psychologist Daniel Kahneman and artificial intelligence pioneer Herb Simon, both Nobel Prize winners. Each embraced a model of a human mind with two parts: a *habitual,* automatic, and largely unconscious mind, along with an *attentive*, reasoning and largely conscious mind. It is likely that the habitual mind represents an older system and is similar to the mental capabilities of early humans: fast, good at complex trade-offs and associations, but not very adept at what we think of as abstract reasoning. Correspondingly, the communication abilities of this early human mind would likely be limited to signaling and simple signs. Although this habitual mind is quite capable of learning new behaviors through experience or mimicry, such learning is likely limited to associations among perceptual features.

The ability to go beyond association-based learning may be the key contribution that the attentive mind makes to the fitness of our species. There are inherent limitations on what associational mechanisms learn and Kahneman has speculated that these limitations probably spurred the evolution of the attentive mind. In addition, the linguistic capabilities of the attentive mind can allow much faster spreading of new behaviors among a population.

OF KITH AND KIN

One of the surprising conclusions from our studies of social signaling in everyday situations is that the attitudes and actions of peers, rather than logic or argument, often dominate people's beliefs and actions. It seems that our forebears understood this intuitively and even had a name for it: *kith*. “Kith and kin” is a thousand-year-old phrase that still rings familiar, but most of us don't know the meaning of “kith.” The word derives from the old English and old German word for knowledge, and it means a more or less cohesive group with common beliefs and customs. These are also the roots for “couth,” which means to act with a high degree of sophistication, as well as its more familiar counterpart, uncouth. Thus, your *kith* is the circle of peers (not just friends) from whom you learn sophisticated habits of action.

It seems that 1,000 years ago the English had the right idea about how people learn. We are ruled by common sense, the habits our kithmates have in *common*. This social learning works by modifying us through social pressure (usually mediated by social signaling), instead of through critical reasoning. The use of kithmates to form “common sense” habits of action is another clue to how early humans may have leveraged the social signaling mechanisms to make better decisions.

It is possible that allowing for more time around the water cooler or coffee pot may be the simplest way to increase workers' productivity. Why? In our studies of more than a dozen organizations, we have found that cohesion among peer employees —kithmates—is one of the largest factors in both productivity and job satisfaction. In these instances, cohesion is defined as how connected kithmates are with each other. That is, do the people you talk to also talk to one another? How tightly woven and interconnected is your peer network?

In one study in Chicago, we used electronic badges to monitor the social signaling and conversational patterns of information technology. The badges were fitted with infrared sensors, Bluetooth location measurement and accelerometers to measure body movement, and recorders that captured the pitch and pace of voices. We found that peer-group cohesion was a central predictor of productivity. In fact, workers whose group cohesion was in the top third had more than 10% higher productivity when compared to the mean. This result underscores the extent to which we are social animals and that our connection with our peers at a local level is vitally important. With increased cohesion comes an increase in things such as shared tacit knowledge, attitudes and work habits, and social support. In other words, much of the important information about how to be successful and productive at a job is likely to be found around the water cooler.

TAPPING COLLECTIVE INTELLIGENCE

But are people always confined by *common* sense—that is, the beliefs of those around them? To answer that, it is important to understand how social signaling mechanisms help people decide when to be guided by kithmates and when to follow a separate path. From a theoretical point of view, perhaps the simplest, most effective way to integrate common sense into people's actions is through an *idea market*. Idea markets resemble voting but instead of building on single votes per person, we allow people to express their expectation of the returns associated with multiple courses of action. For instance, how much food will we find if we go over the hill? How much will we find if we go across the river? And so forth for each alternative. One can think of these expectations as bets and use standard probability theory to weight the bets in proportion to their expected payoff. In this way we can select the action that maximizes the expected return and minimizes the risk.

It is easy to create idea markets using social signaling. Everyone bets on each suggested action by signaling a level of interest. Then group members "add up" the signaling to pick the option with the most positive signaling. This method of decision-making doesn't require language. In order to pick the winning course of action, each participant must only signal to the rest of the group how interested they are in each alternative and then be able to read the group's combined signaling. Animal behavior research supports the idea that this is what both bees and ape troops do when deciding about group movements. It also is similar to the initial reaction signaling seen in business meetings. Those "ums," "ahs" and "hmms" so common in conference rooms, along with the animated or slack body postures, suggest how our modern decision- making processes preserve and leverage these ancient mechanisms.

Still, the challenges individuals face change dramatically over time. As a consequence, social signaling mechanisms must be able to quickly select the right kithmates to help solve the newest problems. This really is a question of identity: The character of the problem determines who will be the best kithmates for learning new, effective actions. This poses a problem for decision making by social signaling, however, because when people are faced with new important decisions, they need to quickly form peer groups that are relevant to the problem. Thus, we need to determine whether or not people dynamically form problem-defined kith groupings in modern daily life.

To test this idea we monitored the social signaling and patterns of interaction for 81 residents in an MIT dormitory during the 2008 presidential elections, giving them smart phones that could track who spoke face-to-face with whom. What we found is that when politics became especially prominent, as during a presidential debate, the students shifted their groupings and began to selectively spend time with others who shared the same ideological position, excluding those holding opposing opinions. This was not true of more remote channels of communication such as phone calls; those remained unchanged, perhaps because they are less effective at conveying social signals.

Further, the extent to which students formed a cohesive kith with people with similar opinions predicted their level of interest in the presidential race, their liberal-conservative balance and even their eventual voting behavior. For first-year students, the magnitude of this effect was similar to the effects detected in other experiments evaluating political advertising and media exposure. The finding reinforces the view that when people are faced with important decisions, they do tend to form into cohesive, reinforcing peer groups, providing the social context and reinforcement for their choices.

WHENCE CREATIVITY?

We have seen that these signals have a major effect on person-to-person interactions and on group behavior, but do they influence even our most sophisticated abilities? As it turns out, the humble honeybee has much to tell us about the flow of information in social species. The notion that worker bees search for good food sources and then return to the hive and use waggle dance signaling to communicate the distance and direction of the food source is common knowledge. Less well known, though, is that bees use this same mechanism as the basis for an elegant approach to group decision-making.

One of the most important group decisions made by a bee colony is where to locate a nest. Bees seem to use a kind of "idea market" to guide their discovery: the colony sends out a small number of scouts to survey the environment. Returning scouts who have found promising sites signal their discovery with an intense, active dance. As a result of this social signaling, more scouts are recruited to the better sites. This cycle of exploration and social signaling continues until, eventually, so many scouts are signaling in favor of the best site that a tipping point is reached and the hive moves en masse.[1]

The bees' decision-making process highlights information integration as well as information discovery, two processes crucial to every organization, but each with different requirements. The solution suggested by the bees is to alternate between the multiple networks that are best for discovery and the richly connected single network that is best

[1] See, for example, "Group Decision Making in Honey Bee Swarms," *American Scientist*, May–June 2006.

for integration. Networks—whether apian or human—that vary their communication structure as needed are able to shape information flow to optimize both discovery and integration.

Our studies at MIT have shown that this same sort of oscillation between discovery and integration seems to be characteristic of creative teams of people. In one study we tracked employees in the marketing division of a German bank, capturing information about their social signaling during each encounter. Analysis of the data showed that teams charged with creating new marketing campaigns oscillated between two communication patterns. In one they placed themselves in the middle of multiple streams of communication, what we call a centralized communication pattern that is associated with discovery. In the second, they engaged in a densely interconnected pattern of communication where most conversations were with other team members. In contrast, members of production groups showed little oscillation, speaking almost entirely to other team members. A second study demonstrated that creative teams not only had more variation in the shape of their social communication network, but also that the range of variation in network shape correlated with how creatively productive the groups judged themselves to be. In other words, oscillation in the shape of these networks can predict creative productivity, at least as defined by the people in the networks.

Why might this pattern promote greater creative output? One way to interpret these findings is that this pattern of oscillation brings new information to a group for integration into our habitual minds. Because the habitual mind uses association rather than logic, it can more easily make intuitive leaps and find new, creative analogies. It can take the experience of a new situation, let it "soak in" for a while and then produce an array of analogous actions. There is considerable literature showing that unconscious cognition is more effective than conscious cognition for complex problem solving. The habitual mind seems to work best when the more logical attentive mind isn't interfering, such as during sleep or when we are "turning it over in the back of our mind.' In contrast, the attentive mind provides insights into our actions, helping us detect problems and work though new plans of action.

> "We can understand at least a pedestrian sort of charisma if we define it by its operational characteristics: an unusual ability to convince others to try out a new behavior"

THE POWER OF CHARISMA

Although using social signaling mechanisms for making decisions appears to be good for combining action alternatives and interests, it is likely not to be good for learning new behaviors. This is because the "idea market" combination mechanism tends to select only consensus views and is unfriendly to new or unusual alternatives. It leads to a very stable, conservative social group. This resistance to change raises the important question of how social signaling mechanisms might have facilitated learning of new action habits from examples outside the community.

One possible mechanism is the phenomenon of charisma. Although no one has fully defined charisma, research subjects reliably agree on its characteristics. In particular, most report that charisma is much more than just word choice or argument. We can understand at least a pedestrian sort of charisma if we define it by its operational characteristics: an unusual ability to convince others to try out a new behavior. Under this definition, people who are good at pitching business plans, building high-performance teams and succeeding at similar activities demonstrate the quality. Importantly, many of these charisma-like effects seem to involve social signaling. In our studies, we have observed that there is a certain style of social interaction—one that we can identify quantitatively and automatically by computer processing of voice and gesture—that is highly predictive of success at influencing others' behavior in a variety of situations.

To illustrate, consider our study on business-plan pitches. In that study, a group of rising-star business executives gathered at MIT for an important task. Each executive presented a business plan to the group, and the group then chose the best ideas. The executives wore our badges, which captured their styles of social signaling. By analyzing that signaling, we were able to predict with a high degree of accuracy which business plans the executives would choose. Our executives, it seems, were busy measuring the social content of the presentations, quite apart from the spoken, informational part.

To understand why this makes sense, consider the situation in more detail. Imagine you are listening to a business plan pitch on an unfamiliar topic. Although you don't know much about the subject, the speaker's presentation is fluid and practiced. Also, the speaker is noticeably energetic and clearly excited. Your habitual mind says to itself, "Well, I may not know much about this, but she is clearly expert and she is excited ... so I guess it must be a good plan." This successful presentation style is charismatic by our definition because it is effective at convincing people to consider new behaviors.

Similarly, another recent study from our research group focused on executives attending a one-week intensive executive education class at MIT where the final project, again, was pitching a business plan. This time we used our electronic badges to observe the executives during a mixer on the first evening of the course. And we found that their social styles at the mixer were predictive of how well their teams' business plans would be perceived at the end of the course. The most successful style is what we call the "charismatic connector". These people circulated in the crowd, practiced intense listening, had fluid speaking styles and tended to drive conversations with questions.

The more charismatic connectors a given team had among its members, the better the team was judged during the business-plan pitch. The reason seemed to be that the members worked together better. In teams whose social style is dominated by these charismatic connectors, team discussions were characterized by more even-handed turn-taking, high levels of engagement and higher productivity. These two characteristics—charisma and connector—usually go together. We have found that the people who have the most consistent and influential style of speaking are also the people who are the greatest connectors. People whose social networks cross many different groups are exactly those people who display a charismatic style of interaction.

UNDER THE SIGNALING INFLUENCE

Our research suggests that people's behavior is much more a function of their social network than generally imagined. Humans truly are social animals, and individuals are best likened to musicians in a jazz quartet, forming a web of unconscious reactions tuned to exactly complement the others in the group. These various studies from my research group all serve to demonstrate that this immersion of self in the surrounding social network is the typical human condition, rather than an isolated example found in exceptional circumstances. Our ancient reflexes for unconscious social coordination fuse us together into problem-oriented peer groups—our kith. And those groups strongly influence our actions every day.

What practical conclusions can we draw from this? These results tell us that individuals should consciously work toward having a cohesive, engaged set of kithmates, helping them adopt more effective habits of action. There is solid evidence that people with cohesive and engaged kithmates are not just more productive and creative but that they are also happier, more resilient and more satisfied. And how can one go about collecting this set of valuable mates? The charismatic connector style of signaling we have uncovered may be the single most important factor in promoting the success of group activities, by creating a contagious positive mood, increasing trust and encouraging more even, socially aware participation. It may be time to begin training people to become more like these connectors.

Reality mining offers insights this promising because its large datasets reveal social patterns that once were invisible. And they can embed, in real-time, pictures of hundreds, even thousands, of people working together. Of course this method raises ethical issues that must be addressed. Such data also pose a potential threat to individual privacy. Because of that, it is important that individuals rather than corporations own data resulting from reality mining. To my mind, that would place control of their use with the observed individuals, where it belongs. And it would also allow the owners to derive personal value from the data. That would create a fair market for public use of such an important knowledge source as we strive to understand how it really is that we work together.

ACKNOWLEDGEMENTS

I would like to thank the students and faculty who collaborated in the research discussed in this article. Tanzeem Choudhury, Taemie Kim, Daniel Olguin Olguin and Ben Wabe helped create the sociometric badges and conduct the related experiments. Nathan Eagle, Anmol Madan and Harvard University Professor David Lazer helped create conduct the smart phone experiments. MIT Assistant Professor Sinan Aral, MIT Professor Erik Brynjolfsson, Tracy Heibec and Lynn Wu helped with the analysis of the experiments. Research papers discussing the resulting research can be found at http://hd.media.mit.edu. An earlier version of this paper was published in American Scientist 98, pp. 204-210.

REFERENCES

Ambady, N., and R. Rosenthal (1992), "Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis", *Psychological Bulletin* 111, pp. 256–274.

Barsade, S. (2002), "The Ripple Effect: Emotional Contagion and its Influence on Group Behavior", *Administrative Science Quarterly* 47, pp. 644–675.

Buchanan, M. (2007), "The Science of Subtle Signals", *strategy+business* 48, pp. 1–10.

Buchanan, M. (2009), "Secret Signals. Does a Primitive, Non-Linguistic Type of Communication Drive People's Interactions?", *Nature* 457, pp. 528–530.

Couzin, I. D. (2009), "Collective Cognition in Animal Groups", *Trends in Cognitive Sciences* 13, pp. 36–43.

Iacoboni, M., and J. C. Mazziotta (2007), "Mirror Neuron System: Basic Findings and Clinical Applications", *Annals of Neurology* 62, pp. 213–218.

Pentland, A., and T. Heibec (2008), *Honest Signals: How They Shape Our World,* Cambridge, MA, The MIT Press.

Zahavi, A., and A. Zahavi (1997), *The Handicap Principle: a Missing Piece of Darwin's Puzzle*, Oxford: Oxford University Press.

BBVA

The Archaeology of Innovation: lessons for our times

Sander van der Leeuw
Arizona State University and
Santa Fe Institute

INTRODUCTION

In this paper, I have tried to bring together a number of strands of work carried out over the last thirty years, both as an archaeologist and as a generalist social scientist, concerned with the very long-term history of human evolution and some of its implications for the challenges of the 21st century. The result is a very personal perspective that notably differs from the contributions of many colleagues in that I have from the outset posited that what characterizes modern human (*Homo sapiens sapiens*)[1] behavior and modern human societies is information processing that includes learning and learning how to learn (second-order learning, see Bateson 1972), as well as categorization, abstraction, (hierarchical) organization and related phenomena. Moreover modern humans communicate between themselves by various kinds of symbolic means, and have the capacity to transform their natural and material environment in many different ways, and at many spatial and temporal scales. As a result, this paper does diverge from the usual population-based Darwinian thinking about human evolution (*e.g.* Boyd and Richerson, 1985 *etc.*) in that, for the later periods (*cf.* Lane *et al.,* 2009), it focuses on 'organization thinking'—studying the evolution of the ways in which human beings process information, organize themselves, and transform the world around them.

Necessarily, this paper takes the shape of an introductory summary of many of the underlying arguments about the trajectory of human evolution and the aspects of that history that are particularly relevant to the present and the future. Where possible, I have referred to papers and other publications that elaborate my main train of thought. However, I have kept other references to a minimum, not wanting to load the argument with the many doubts and discussions that have occurred in the anthropological and archaeological community over the period of gestation. I have thus been able to reserve space to point out some of the implications of this approach for present-day challenges, in particular the contradiction between two of today's favorite buzzwords: 'innovation' and 'sustainability'.

The evolutionary history of the human species, and in particular its cognitive and organizational capacity, is here seen as consisting of two parts, the first of which is essentially biological (the growth of our brain

[1] The distinction between humans (*Homo sapiens*) and modern humans (*Homo sapiens sapiens*) referred to here follows current custom among paleo-anthropologists. The transition is estimated to have occurred somewhere around 200,000 years BP.

and its cognitive capacity), whilst the second is essentially cultural (learning to exploit the full capacity of the brain). Hence, this paper is divided into three major sections, describing respectively 1. the biological evolution, 2. the cultural evolution and 3. the implications of the species' past history for our present-day challenges.

It should be emphasized that each of these three sections is based on insights and knowledge from different disciplines and sub-disciplines. The first part derives from arguments in evolutionary biology and evolutionary psychology, and therefore is based on an essentially life-science epistemology and argument, and data deriving from ethology, palaeo–anthropology and cognitive science. It attempts to reconstruct the evolution of the human species leading up to its present-day capabilities by comparing living primates, the fossil remains of—and the artifacts made by—humans at various stages of their development, and the physical and behavioral characteristics of modern human beings. This leads to a patchwork of data-points and ideas that, in so far as it coherently holds together, derives its principal interest from the fact that it raises new questions and provides a basis for the arguments in the second part.

That part, on the other hand, derives from arguments in archaeology and history, which are based on humanities and social science epistemologies respectively, and data and insights from archaeological, written historical and modern observational sources. It attempts to outline the development of societal organization from small roaming gatherer-hunter-fisher bands, via villages, urban systems and empires to the present day global society, with a focus on the role and forms energy and information play in that development. In doing so, I am using the constraints and opportunities afforded by the bio-social nature of our species to explain observed phenomena in human history, and couching the explanation in systemic terms, which many archaeologists and most historians would have difficulty recognizing. And to add insult to injury, I am doing so at a level of generalization that is beyond any commonly used in these disciplines.

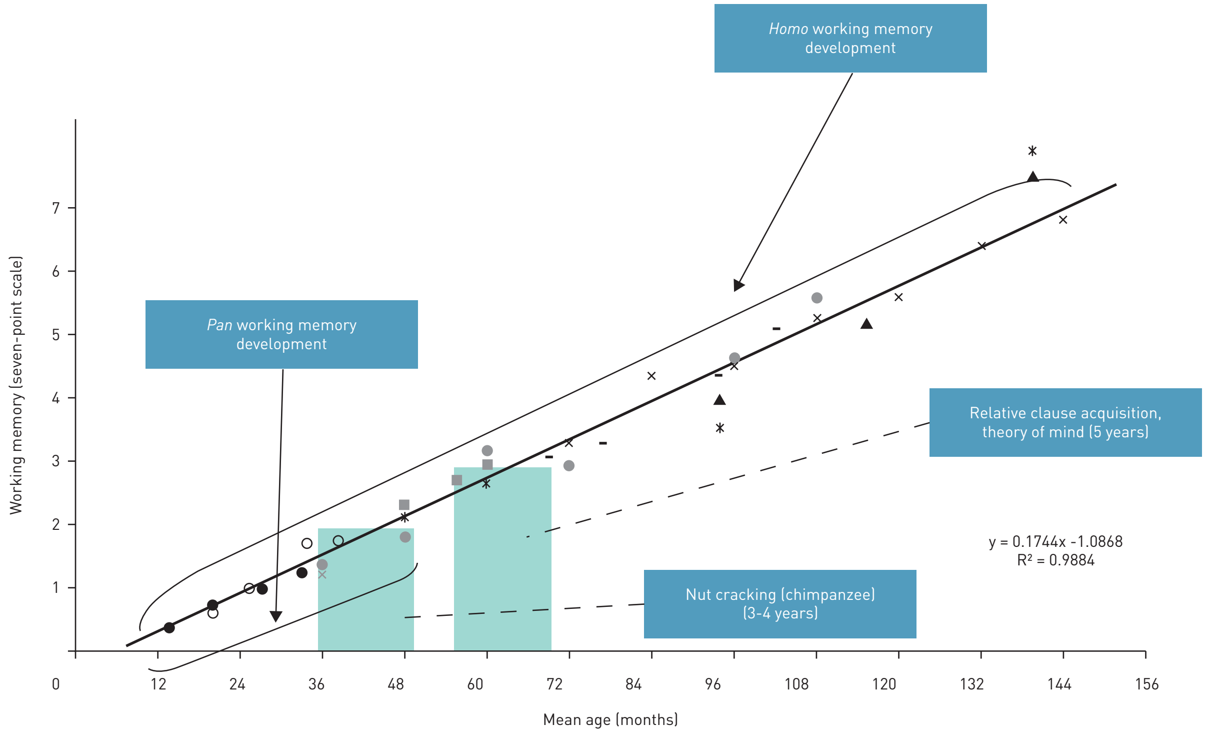
My justification for doing this is the fact that most, if not all, trans-disciplinary research must aim to "constructively upset the practitioners of all the disciplines involved" in order to raise new questions and challenges to be considered by the communities practicing these disciplines as well as by others, and thus to 'stretch the envelope' of our knowledge and insights. The direction in which I have attempted to stretch that envelope is given by the fact that this paper intends to make a contribution to the current sustainability debate.

In the third part of the paper, I have tried to outline how the bio-social nature of human beings and the course of the history of the species over the last 12,000-15,000 years have conspired to create the dilemma that we face today: "How do we use the human capacity to innovate, the unbridled use of which during the last three centuries has caused the unsustainability of our current mode of life, to attain a more sustainable society?" The short answer is clearly: "We must use our capacity to innovate in a different way!" This third part of the paper therefore ends with some suggestions derived from observing a fundamental weakness of our current scientific thinking—the capacity to derive lessons from the past for the future.

THE EVOLUTION OF THE HUMAN BRAIN

The first part of the story concerns the physical development of the human brain and its capacity to deal with an increasing number of simultaneous information sources. The core concept that is most relevant here is the evolution of the short-term working memory (hereafter STWM), which determines how many different sources of information can be processed together in order to follow a particular train of thought or course of action. There are different ways to reconstruct this evolution (Read and van der Leeuw, 2008, 2009). Indirectly, it can be interpolated by comparing the STWM of chimpanzees (our closest common ancestor in the evolutionary tree that produced modern humans) to that of modern human beings. 75% of chimpanzees are able to combine three elements (an anvil, a nut and a hammerstone) in the act of cracking the nut, which leads us to think that the STWM of chimpanzees is 2 ± 1 (because 25% of them never master this). Experiments with different ways of calculating the human capacity to combine information sources, on the other hand, seem to point to an STWM of 7 ± 2 for modern humans. This difference coincides nicely with the fact that chimpanzees reach adolescence after 3-4 years, and modern humans at age 13-14. It is therefore assumed that the growth of STWM occurs before adolescence in both species, and that the difference in age of adolescence explains the difference in STWM capacity (Figure 1, *cf.* Read and van der Leeuw, 2008: 1960).

Figure 1. The relationship between cognitive capacity and infant growth in *Pan* and in *Homo sapiens sapiens*. The trend line is projected from the regression of time-delay response (Diamond and Doar, 1989) on infant age. Data are rescaled for each dataset to make the trend line pass through the mean of that dataset. Working memory scaled to STWM = 7 at 144 months. The 'Fuzzy' vertical bars compare the age of nut cracking among chimpanzees with the age for relative clause acquisition and theory of mind conceptualization in humans. (Data on STWM are here represented by the following symbols: • = Imitation (Alp 1994); + = time delay (Diamond and Doar, 1989); ✳ = number recall (Siegel and Ryan 1989); x = total language score (Johnson *et al.* 1989); x = relative clauses (Corrêa, 1995; ■ = count label, span (Carlson *et al.*, 2002); o = 6 month retest (Alp, 1989); ▲ = world recall (Siegel and Ryan, 1989); ● = spatial recall (Kemps *et al.*, 2000); ♦ = relative clauses (Kidd and Bavin, 2002); - = spatial working memory (Luciana and Nelson, 1998); — =linear time delay (Diamond and Doar, 1989))



Another approach to corroborating the growth of STWM is by measuring encephalization—the evolution of the brain-to-body-weight ratio of modern humans' ancestors through time. The evolution of these ratios is based on the skeletal remains of each subspecies found and, as shown in Figure 3, corresponds nicely to the evolution of the STWM as has been established based on the way and extent to which these ancestors were able to shape stone tools (*cf.* Read and van der Leeuw, 2008: 164).

Whereas both these approaches depend in fact on extrapolation and therefore do not provide any *direct* proof for our thesis, the study of the way and extent to which the various subspecies and variants preceding modern humans have been able to shape stone tools does provide some direct evidence, which is summarized in Table 1. That table links the evolution of actions in stone tool making with the concepts that they define, the number of dimensions and the STWM involved with the stone tools that provide examples of each stage.

In order to explain the development involved, I will use an example: the mastering of three-dimensional conceptualization and manufacture of stone tools (*cf*. figure 2 a-d) (Pigeot, 1991; van der Leeuw, 2000). The first tools are essentially pebbles from which at one point of the circumference (generally where the pebble is pointed) a chip has been removed to create a sharper edge (fig. 2a). Removing the flake requires three pieces of information: the future tool from which the flake is removed, the hammerstone with which that is done, and the need to maintain the two at an angle of less than 90º at the time of the blow. Here, we therefore have to do with proof of STWM 3. In the next stage, this action (flaking) is repeated along the edge of the pebble. That requires control over the above three variables, and a fourth one: the succession of the blows in a line. STWM is therefore 4 (figure 2b). Next, the edge is closed: the toolmaker goes all around the pebble until the last flake is adjacent to the first. By itself, this is not a complete new stage, and we have called this STWM 4.5. But once the closed loop is conceived as defining a surface the knapper has two options. Either to define a surface by knapping an edge around it and then taking off the centre, or to do the reverse—take off the centre first, and then refine the edge. The conceptual reversibility shows that the knapper has now integrated five dimensions, and that his or her STWM is 5 (figure 2c). The next stage again develops sequentiality, but in a more complex way. In the so-called 'Levallois' technique, making one artifact serves at the same time as preparation for the next, by dividing the pebble conceptually in two parts along its edge. And finally, the knapper works completely in three dimensions, preparing two surfaces and then taking flakes off the third. At this stage, STWM 7 (figure 2d), for the first time the knappers are able not only to work a three-dimensional piece of stone, but also to conceive it as three-dimensional and adapt their working techniques accordingly, greatly reducing loss and increasing efficiency.

Closely observing the tools and other traces of human existence available around 50,000 BP indicates that, after some 2,000,000 years, people at that time could (van der Leeuw, 2000):

- distinguish between reality and conception;
- categorize based on similarities and differences;
- in their thinking, feed-back, feed-forward and reverse in time (*e.g.* reverse an observed causal sequence, in order to

conclude from the result what kind of action could achieve it);

- remember and represent sequences of actions, including control loops, and conceive of such sequences that could be inserted as alternatives in manufacturing sequences;
- create basic hierarchies, such as point-line-surface-volume, or hierarchies of size or inclusion;
- conceive of relationships between a whole and its constituent parts (including reversing these relationships);
- maintain complex sequences of actions in the mind, such as between different stages of a production process;
- represent an object in a reduced set of dimensions (*e.g.* life-like cave paintings).

Figure 2. Graph of encephalization quotient (EQ) estimates based on hominid fossils and *Pan* (Chimpanzees). Early hominid fossils have been identified by taxon. Each data point is the mean for hominid fossils at that time period. Height of the 'fuzzy' vertical bars is the hominid EQ corresponding to the data for the appearance of the stage represented by the fuzzy bar. Right vertical axis represents STWM. Data are adapted from the following: triangles: Epstein 2002; squares: Rightmire, 2004; diamonds: Ruff *et al.*, 2004. EQ= brain mass/(11.22 body mass$^{0.76}$), *cf.* Martin, 1981.

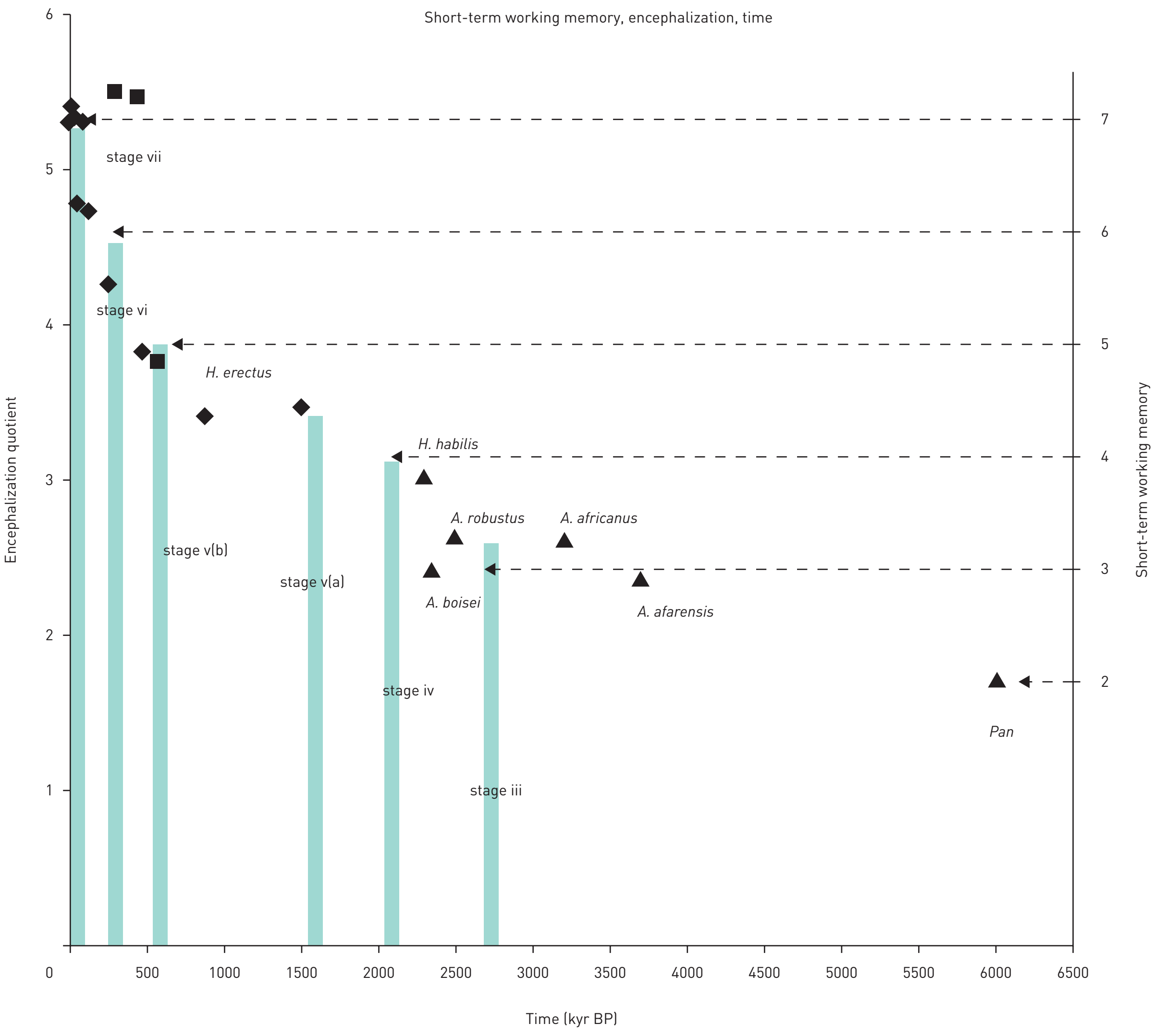
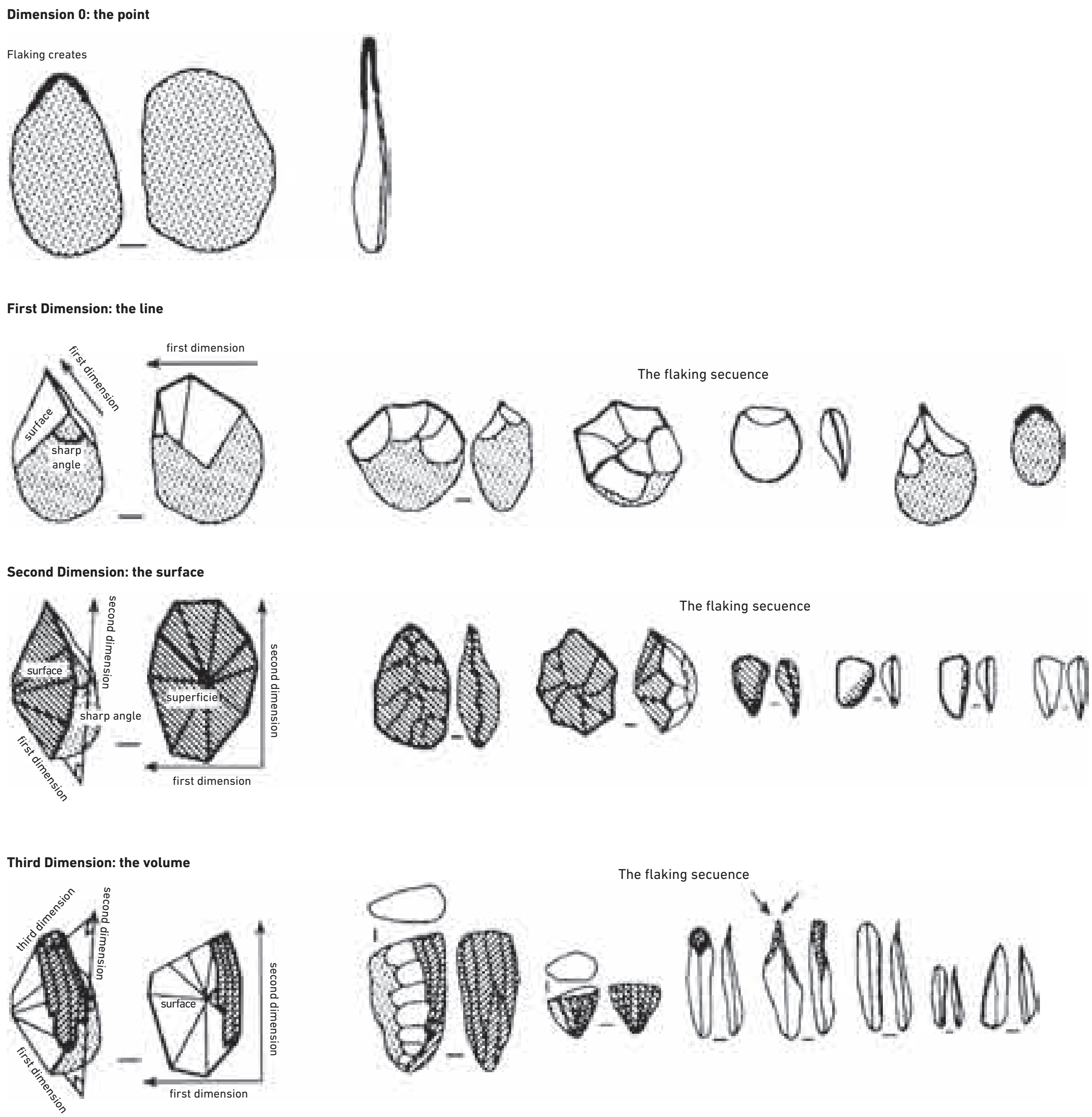


Table 1. Evolution of stone tool manufacture from the earliest tools (stage 2, " 2,6 M. years ago; found in Lokalalei 1) to the complex blade technologies (stage 7, found in most parts of the world c. 50,000 BP). Columns 2-5 indicate the observations leading us to assume specific STWM capacities; Column 8 (bold) indicates the stage's STWM capacity and column 9 the approximate age of the beginning of each stage. Column 10 refers to the relevant artifact categories documenting the stages. For a more extensive explanation, see Read and van der Leeuw, 2008: 1961-1964).

Stage	Concept	Action	Novelty	Dimensionality	Goal	Mode	ST WM	Age BP	Example
1	Object attribute	Repetition possible	Functional attributes already present; can be enhanced	0	Use object		**1**		
1A	Rela-tionship between objects		Using more than one object to fulfill task	0	Combine objects		**2**		
2	Imposed attribute	Repetition possible	Object modified to fulfill task	0	Improve object		**2**	> 2,6 My	Lokalalei 1
3	Flaking	Repetition	Deliberate flaking, but without overall design	0: Incident angle < 90°	Shape flakes		**3**	2,6 My	Lokalalei 2C
4	Edge	Iteration: each flake controls the netxt	*Débitage*: flaking to create ad edge on a core	1: Line of flakes creates partial boundary	Shape core	1	**4**	2,0 MY	Oldowan chopper
5	Closed curve	Iteration: each flake controls the next	*Débitage*: flaking to create an edge and a surface	2: Edges as generative elements of surfaces	Shape biface from edge	2	**4,5**		
5A	Surface	Iteration: each flase controls the netx	*Façonnage*: flaking used to make a shape	2: Sufraces intended elements, organized in relation to one another	Shape biface from surfaces	2	**5**	500 Ky	Biface hand-axes
6	Surface	Algorithm: removal of flake prepares next	Control over location and angle of flaking to form surface	2: Surface of the flake brought under control but shape constraint	Serial production of tools	3	**6**	300 Ky	*Levallois*
7	Intersection of planes	Recursive application of algorithm	Prismatic blade technology; monotonous process	3: flake removal retains core shape—no more shape constraint	Serial production of tools	4	**7**	> 50 Ky	Blace technologies

Figure 3. For humans to attain the capacity to conceive of a three-dimensional object (a pebble or stone tool) in three dimensions takes around 2 million years. a. Taking a flake off at the tip of the pebble is an action in 0 dimensions, and takes STWM 3; b. successively taking off several adjacent flakes creates a (1-dimensional) line, and requires STWM 4; c. stretching the line until it meets itself defines a surface by drawing a line around it and represents STWM 4.5; distinguishing between that line and the surface it encloses implies fully working in two dimensions, and requires STWM 5; c. preparing two sides in order to remove the flakes from the third side testifies to a three-dimensional conceptualization of the pebble, and requires STWM 7. (From: van der Leeuw, 2000)



THE INNOVATION EXPLOSION: MASTERING MATTER AND LEARNING HOW TO PUT THE BRAIN TO BEST USE

After 50,000 BP[2], and especially after around 15,000 BP, we see a true 'innovation explosion' occurring just about everywhere on Earth. The sheer multitude of inventions in every domain is truly astonishing, and accelerates up to the present day. There is no reason to assume further developments of the human STWM, as the experimental evidence indicates that modern humans currently have the capacity to deal simultaneously with at most seven, eight or sometimes nine dimensions or sources of information, but even a superficial scrutiny of modern technologies, languages and other achievements shows the wide variety of things that can be achieved with a STWM of 7±2. We would therefore argue that for this next phase, *from about 50,000 BP to the present, the biology of the mind does no longer impose any constraints, and the emphasis is on acquiring the fullest possible range of techniques exploiting the STWM capacity available.*

The emergence of improved technologies

We can distinguish several phases in that process. In the first, the toolkit explodes, but the gatherer-hunter-fisher mobile lifestyle remains the same. Some of the many cognitive operators that emerge in that first stage are (van der Leeuw, 2000):

- the use of completely *new topologies* (e.g. that of a solid around a void, such as in the case of a pot or basket);
- the use of many *new materials* to make tools with. Although it is difficult to prove that these materials were not used earlier, nevertheless, one observes from this time onwards objects in bone, as well as wood and other perishable materials;
- the *combination of different materials into one and the same tool* (e.g. hafting small sharpened stone tools into a wooden or bone handle);
- the *inversion of the manufacturing sequence from reductive* (one that begins with a big object (a block of stone) and successively takes smaller and smaller pieces off it) to gain control over the shape, *to additive* (where tiny particles (clay, fibers) are combined into larger, linear objects (threads, coils) and then into a two-dimensional object (such as a woven cloth), that is then given shape (by sewing) to fit a three-dimensional object (a piece of clothing), etc. This implies the *cognition of a wide range of scales*;
- *stretching and chunking the sequence of actions* kept in the mind: distinguishing between (complex) preparation stages (e.g. gathering of raw materials, preparing them, shaping of pottery, drying, decorating, firing) yet being able to link the logic of manufacture across these stages (adapt the clay to the firing technique, etc.),

The resulting explosion of new tools characterizes the period until about 13,000 BP (in East Asia) or 10,000 BP (in the Near East). But the subsistence mode was still characterized by a *multi-resource strategy* of *harvesting* various foodstuffs in the environment, but now including a wider range facilitated by the new toolkit, adapting to change (weather, availability of food) by *moving around*, albeit over increasingly limited distances, so as to always *stay below the carrying capacity* of the environment. In effect, people lacked the know-how to *inter-act* with their environment; they could only *re-act* to it. *Uncontrollable change and risk* were the order of the day, but people did minimize risk where they could (*cf.* van der Leeuw, 2000).

[2] All the dates mentioned in this paper are not only approximate, and differ between different parts of the world, but are also continually subject to revisions as archaeological research progresses.

The first villages, agriculture and herding

In the next stage, *c.* 13,000-10,000 BP, the continued innovation explosion changed the whole lifestyle of many humans. The acceleration was so overwhelming that in a few thousand years the whole way of life of most humans on earth changed: rather than live in small groups that roamed around, people concentrated their activities in smaller territories, invented different subsistence strategies, and in some cases literally settled down in small villages (van der Leeuw, 2000, 2007 and references therein).

Together, these advances greatly increased the number of ways at people's disposal to tackle the challenges posed by their environment. That rapidly increased our species' capability to invent and innovate in many different domains, allowed it to meet more and more complex challenges in shorter and shorter timeframes, and thus substantively increased humans' adaptive capacity. But the other side of the coin was that these solutions, by engaging people in the manipulation of a material world that they now partly controlled, ultimately led to new, often unexpected, challenges that required the mobilization of great effort in order to overcome them in due time.

As part of this process, a number of fundamental changes occurred. First of all, the relationship between societies and their environments became *reciprocal*: the terrestrial environment from now on did not only impact on society, but society also impacted on the terrestrial environment. As a result, sedentary societies tried to *control environmental risk* by *intervening in the environment*, notably by: 1. *narrowing and optimizing the range of their dependencies on the environment;* 2. *simplifying or even homogenizing (parts of) their environments;* and 3. *spatial and technical diversification and specialization (cf.* van der Leeuw, 2000). The new subsistence techniques introduced, including horticulture, agriculture and herding, narrowed the range of things people depended on for their subsistence. In the process, certain areas of the environment were 'cleared' and dedicated to the specific purpose of growing certain kinds of plants. This required *investment* in certain parts of the environment, devoting those areas to specific activities and delaying the rewards of the investment activities. Clearing the forest and sowing resulted only a year later in a harvest, for example.

The resulting increase of investment in the environment in turn anchored different communities more and more closely to the territory in which they chose to live. People now built permanent dwellings *using the new topology* (upside down containers), and devised many other new kinds of tools and tool–making technologies facilitating the new subsistence strategies practicable in their environment (*e.g.* the ard, the domestication of animals, baskets and pottery for storage, pottery for boiling). Without speaking of (full-time) 'specialists', certain people in a village began to dedicate more time, for example, to weaving or pottery-making, and in doing so provided the products of their work to others in exchange for some of the products these others produced. Differences in resource availability and technological know-how thus led to economic diversification and, in order to provide everyone with the things they needed, the emergence of trade.

The symbiosis that thus emerged between different landscapes and the ways invented and constructed by human groups to deal with them, *by narrowing the spectrum of adaptive options open to the individual societies concerned, drove each of them to devise more and more complex solutions, with more and* *more unexpected consequences that then needed to be dealt with in turn.*

In keeping with my fundamental tenet that information processing is crucial to such changes, I attribute the changes outlined in this section to the beginnings of a new dynamics, in which *learning moved from the individual to the group* because the dimensionality of the challenges to be met increased beyond the capability of individuals to deal with them. This involved the emergence of the following feedback loop (van der Leeuw, 2007):

> Problem-solving structures knowledge —→ more knowledge increases the information processing capacity —→ that in turn allows the cognition of new problems —→ creates new knowledge —→ knowledge creation involves more and more people in processing information —→ increases the size of the group involved and its degree of aggregation —→ creates more problems —→ increases need for problem-solving —→ problem-solving structures more knowledge ... etc.

It enabled the continued accumulation of knowledge, and thus of information-processing capacity, which in turn enabled a concomitant increase in matter, energy and information flows through the society, and thus the *growth of interactive groups*. But this growth was at all times constrained by the amount of information that could be communicated among the members of the group, as miscommunication would have led to misunderstandings and conflicts, and would thus have impaired the cohesion of the communities involved. Communication stress did in my opinion provide the incentive for 1. *improvements in the means of communication* (for example by 'inventing' new, more precise, concepts with which to communicate ideas (*cf.* van der Leeuw, 1982), and 2. *a reduction in the search time needed to find those with which one needed to communicate* (by adopting a sedentary lifestyle).

Finally, as the social system diversified, and people became more dependent on each other, the risk pattern increasingly also included social stresses caused by misunderstandings and miscommunications. Handling risks therefore came to rely increasingly on social skills, and the collective invention and acceptance of organizational and other tools to maintain social cohesion.

The first towns

From this point in time, I will no longer try to point out any novel innovations or cognitive operations emerging as human societies grew in size and towns spread over the surface of the earth. Instead, I will focus on how the feedback system that drove societal growth as well as the conquest of the material world through innovation posed some major challenges. Overcoming these ultimately enabled the emergence of true 'world systems' such as the colonial empires of the early modern period (van der Leeuw, 2007) or the current globalized world.

Throughout the third stage, from around 7,000 BP, *communication remained a major constraint* because more and more people were interactive with each other as the size of settlements involved grew to what we now call a town. This stage—again—sees the emergence of a host of new innovations, such as writing, periodic markets, administration, laws, bureaucracies, specialized full-time communities engaged in specific activities (priests, scribes, soldiers, different kinds of craftsmen and women, etc.). Many of these had either to do with improving communication (such as writing and scribes), social regulation (administration, bureaucracies, laws), the harnessing of more and more resources (mining) or the exchange of objects and materials in part over longer and longer distances (markets, long-distance traders, innovations in transportation). But as larger groups aggregated, the territory ('footprint' to use a modern term) upon which they depended for their material and energy needs expanded exponentially, and the effort required to transport foodstuffs and other materials did the same. *This caused the emergence of energy as a major constraint that did handicap the evolution of societies for millennia to come.*

To deal with this constraint, an interesting core-periphery dynamic emerged to exploit that ever-growing footprint—the exchange of organization against energy. *Around towns, dynamic 'flow structures' emerged in which organizational capacity was generated in the towns and then spread around them*, extending the town's control over a wider and wider territory; in turn, the increasing quantities of energy collected in that territory (in the form of foodstuffs and other natural resources) flowed back towards the city to feed the ever-increasing population that kept the flow structure going by innovation (creation of new organization and information-processing capacity). These 'flow structures' became the 'bootstrapping' drivers that created larger and larger agglomerations of people and the territories to go with them.

What enabled the urban populations to keep innovating, and thus to maintain the flow structures, was—again—the growing capacity of more and more interacting minds to identify new needs, novel functions and new categories, as well as new artifacts and challenges. Underpinning that dynamic is one that we know well in the modern world. Invention is usually (and certainly in prehistoric and early historic times) something that involves either individuals, or very small teams. Hence, in its early stages it is related to a relatively small number of cognitive dimensions—it solves challenges that few people are aware of. As such inventions become the focus of attention of much larger numbers of people, they simultaneously become cognized in many more dimensions (people see more uses for them, ways to slightly improve them, etc.), and this in certain cases triggers an 'innovation cascade'—a string of further innovations, including new artifacts, new uses of existing artifacts, and new forms of behavior and social and institutional organization. In this process, clearly, towns and cities are more successful than rural areas because of the greater number of interactive individuals in such aggregations. That is corroborated by the fact that when scaling a number of urban systems of different sizes against respectively metrics of population, energy and innovation, population scales linearly, energy sub-linearly and innovation capacity super-linearly (Bettencourt *et al.,* 2006)

Empires

The above 'flow structures' kept growing (albeit with ups and downs) until, after several millennia (from about 2500 BC in the Old World, and about 500 BC in the New), they were able to cover very large areas, such as the prehistoric and early historic empires (The Chinese, Achaemenid and Macedonian and Roman Empires, for example, in the Old World, the Maya and Inca Empires in the New World, and later the European colonial Empires), which concentrated large numbers of people at their center (and, in order to feed them, gathered treasure, raw materials, crops and many other commodities from their hinterlands). *Throughout this period communication and energy remained the* *main constraints,* impacting on cities, states and empires. Thus we see advances in the harnessing of animal energy (including slavery), wind power (for transportation in sailing vessels and for driving windmills), falling water (for mills), etc., but also in the facilitation of communication, (e.g. long-distance 'highways' over land, the sextant and compass to facilitate navigation at sea), and in all kinds of ways to create and concentrate wealth serving to defray the costs of managing societal tensions, maintaining an administration and an army, etc.

Those costs effectively limited the extent of Empires in space and time. Tainter (1988), for example, argues convincingly that only the treasure accumulated outside the Roman Empire in the centuries before the Roman conquest enabled Rome to maintain the large armies and bureaucracies to keep its Empire. As soon as there was no more treasure to be gained by conquering, and the Empire was thrown back upon a dependency on recurrent (in essence solar) energy, he argues that it could no longer maintain the flow structure. This reduced the advantages of being part of the Empire, so that it began to lose control over its wide territory, causing people to fall back on smaller, regional or local networks. Thus disaffection, or even dispersion of the population, followed the cessation of the flows that generated the coherent socio-economic structure of an Empire in the first place.

THE LAST THREE CENTURIES

The last three centuries have seen the (provisional) culmination of the trajectory I have outlined in Part II. That trajectory shows how the constraints and opportunities afforded by the bio-social nature of our species explain a number of observed phenomena if human history is conceived in systemic terms. In that sense, these last three centuries do not differ from what went before, but they have seen an unbridled acceleration of our species' innovative activity, initially because the 'taming' of fossil energy removed the energetic constraint from much human activity, and subsequently because the introduction of electronics enabled the separation of information from most of the substrates used for its transmission until then. These two developments together have engendered a 'quantum jump' or 'state change' in societal dynamics, which has been at the root of many of today's challenges, but also introduces potential ways to deal with them that were not available thus far.

The introduction of fossil energy and society's dependence on innovation

The (for the moment) last phase of this long-term process of social evolution through innovation involves the last two and a half centuries, in which first the energy constraint was removed by the introduction of plentiful fossil energy, and recently the communication and information- processing constraint is in the process of being removed due to the development of new technologies. The introduction of fossil energy first brought in its wake new technologies to enable, facilitate or reduce the cost of transportation (railroads steamers, cars, etc.), manufacturing (steam-driven factories), and energy itself, as well as (later) technologies to reduce the amount of energy needed to fulfill societal needs.

Without immediately having a clear explanation, however, I would like to signal another emergent driver that, in this period, transformed innovation from a demand-driven activity to a supply-driven one. For most of human (pre-) history, it seems that inventions were either the result of perceived

"Between the 18th and the 20th centuries, and particularly in the second half of the latter, with respect to innovation, *the balance between supply and demand shifted in favor of supply*. Rather than societal needs driving innovation, innovation came to drive societal needs"

needs, or were not really introduced on a large scale in societies until such a need emerged. It took, for example, roughly 1000 years after the invention of ironworking to actually see that technique spread throughout Europe at a fairly rapid pace (*cf.* Sørensen). In that case, the initial brake on the transformation of this invention into an innovation seems to have been related to the social structure of society. In the Bronze Age, hierarchies emerged that controlled wide exchange networks because they controlled the sources of bronze, which was relatively easy to do because accessible sources to this metal were relatively few and far between. That is not the case with iron—it can be found in virtually every water-rich place in Europe, and once the technology to use it spread, no one could any longer derive riches from controlling the manufacture of iron tools. The introduction of iron technology therefore enabled large numbers of people to manufacture and use much better tools and weapons and had, in a sense, a democratizing effect.

Between the 18th and the 20th centuries, and particularly in the second half of the latter, with respect to innovation, *the balance between supply and demand shifted in favor of supply*. Rather than societal needs driving innovation, innovation came to drive societal needs. Companies competed to lay their hands on inventions (or developed them internally), and then created markets for them, forcing their use on society in order to enhance their profit. *This has led to a situation in which innovation has become endemic to our societies, and those societies, through their dependency on ever-increasing GDP and profit figures, have become dependent on innovation for their continued existence.* This is a novel dynamic that has major consequences for the way we might deal with the challenges of the 21st century, sustainability among them. I will come back to this in a later section.

This phenomenon has emerged in a period that saw the transformation of our society's perspective on time. Whereas until the 17th century, the most frequent vision explained the present by invoking 'History' or 'The Past' or 'It has always been like this', whereas invoking something 'new' or 'an innovation' was socially heavily frowned upon. With the enlightenment this changed, ultimately leading to our current attitude, in which the 'new' is mostly preferred over the 'old', the 'proven' or the 'heritage' (Girard, 1990). Interestingly enough, this change in perspective was accompanied by the institutionalization of the universities and academic disciplines as 'research crucibles', initially on the expectation that, ultimately, something useful would be invented, but increasingly with the expectation that such economic advantages are what research exists for.

Separating information from its material and energetic substrates

Although 'information technology' has been in existence for many thousands of years, in the form of gestures, language, writing, accounting, and many other things including North American smoke signals and African tamtams, the second half of the 20th Century saw the definition of the concept of 'information (Shannon and Weaver, 1948) and rapidly thereafter the mechanization of information processing, initially in the domain of communication, but then also in the domains of calculation, representation and many others. Hence, the current emphasis in certain quarters on our present-day society as the 'information society' is misguided—every society since the beginning of human evolution has been an 'information society'.

Clearly, as we are only at the beginning of a process that will eventually harness electronic and other forms of information processing throughout all aspects of our thinking and our society, and offer many new solutions to existing challenges and equally many new challenges, we cannot presently outline the higher-level 'drivers' that may emerge as a result of that process. However, we do note that, again, these will accelerate the dependency of our society on innovation. Indeed, massive information collection and treatment, as well as the application of the concept of information to physical, biological and societal processes, is emerging as a new challenge: the NBIC 'revolution', under which we understand the encounter (and potential interaction) of nano- bio- information- and communications technology.

However that may be, after the mastering of 'matter' by devising ways to conceptually separate manipulating it from the time/space in which that process occurred, which took humanity about two million years, and the mastering of energy by separating it conceptually from movement and change, which took the next 7000 years, it took only 200 years more to conceptualize information by separating it from its material or energetic substrate. Our collective capability to process information has therefore accelerated more or less exponentially, as has the size of Earth's human population and—more important from our perspective—the size and number of the cities that are the principal source of new inventions and innovations. Having identified the driver behind this process, as with any such exponential growth, we have to ask: "How much longer can this go on?" In order to answer that question, we must look at the long-term consequences of the 'innovation explosion', from the Neolithic to the present.

THE CHALLENGE OF THE FUTURE—INNOVATION, SUSTAINABILITY AND 'UNANTICIPATED CONSEQUENCES'

One way to introduce this topic, to which we will devote the last part of this paper, is to point out the contradiction in the fact that innovation is seen as the way out of the present syndrome of overpopulation, looming or current resource shortage, omnipresent pollution, etc., even though two centuries of unbridled innovation are responsible for bringing about the consumer society as well as the current sustainability challenge. One must conclude that innovation as it is presently embedded in our societies is hardly the panacea to get us out of the sustainability predicament that many claim it is. That in turn prompts the question whether there are any alternatives to 'innovating ourselves out of trouble', and if there are, what could they be?

It seems to me that the root of this challenge lies in the relationship between the fundamental limitations of the human mind, whether collective or individual, and the complexity of the world outside us. I would argue that, over the millennia, that relationship has changed as a result of the innovation explosion itself. In order to understand the nature of that change, we need to look at the relationship between people and their environment.

Human cognition, powerful as it may have become in dealing with the environment, is only one side of the (asymmetric) interaction between people and their environment, the one in which the perception of the multidimensional external world is reduced to a very limited number of dimensions. The other side of that interaction is human action on the environment, and the relationship between cognition and action is exactly what makes the gap between our needs and our capabilities so dramatic. The crucial concept here is that of 'unforeseen' or 'unanticipated' consequences. It refers to the well-known and oft-observed fact that, no matter how careful one is in designing human interventions in the environment, the outcome is never what it was intended to be. It seems to me that this phenomenon is due to the fact that every human action upon the environment modifies the latter in many more ways that its human actors perceive, simply because the dimensionality of the environment is much higher than can be captured by the human mind. In practice, this may be seen to play out in every instance where humans have interacted in a particular way with their environment for a long time—in each such instance, ultimately the environment becomes so degraded from the perspective of the people involved that they either move to another place or change the way they are interacting with the environment.

How does this happen? Imagine a group of people moving into a new environment, about which they possess little knowledge, such as the European settlers into the Eastern North American forests (Cronon, 1983). After a relatively short time, they will observe challenges or opportunities to interact with this environment, and they will 'do something' about them. Their action upon these challenges is based on an impoverished perception of them, which mainly consists of observations concerning the short-term dynamics involved. Yet these same actions transform the environment in ways that affect not only the short-term, but also the long-term dynamics involved in unknown ways. Over time, little by little all the frequent challenges become known and are modified by the society's interaction with the environment, while the unknown longer-term challenges that are introduced accumulate. Or to put this in more abstract terms, due to human interaction with the environment, the 'risk spectrum' of the socio-environmental system is transformed into one in which unknown, long-term (centennial or millennial) risks accumulate to the detriment of shorter-term risks.

Ultimately, this necessarily leads to 'time-bombs' or 'crises' in which so many unknowns emerge that the society risks being overwhelmed by the number of challenges it has to face simultaneously. It will initially deal with this by innovating faster and faster, as our society has done for the last two centuries or so, but as this only accelerates the risk spectrum shift, this ultimately is a battle that no society can win.

There will inevitably come a time when the society drastically needs to change the way it interacts with the environment, or it will lose its coherence. In the latter case, after a time, the whole cycle begins anew—as one observes when looking at the rise and decline of firms, cities, nations, empires or civilizations.

What is the effect of an exponential increase in information-processing capacity on this asymmetry between human understanding and human action? Clearly, as the information-processing capacity increases, the total number of (collectively) cognized dimensions involved in the process does so more or less commensurately. The human actions on the environment therefore affect more and more dimensions of the processes going on in that environment. As the multiplier between cognized human dimensions and unknown environmental dimensions affected by human actions is large, this implies that due to the exponential increase in the number of human cognized dimensions, the number of affected environmental dimensions grows even more rapidly, posing ever more rapidly ever more complex environmental challenges for humankind to deal with.

This permanent, and increasing, tension between the total cognitive capacity of a society and the complexity of its environment has in itself been a, if not the, major driver behind the increase in information-processing capacity of human beings and societies. As such, it has had important consequences for the information-processing structure of the societies involved. Several of these have already been mentioned in this paper: population increase, aggregation of human populations in villages and then cities, the invention of writing, markets, administration and other phenomena accompanying urbanization, etc. But others have not been given much attention; such as its impact on our language and the way we have done (and often still do) science.

Let us look at language first. Initially, as small groups lived together most of the time, humans had the opportunity and time for multi-channel communication—spoken language, gestures, body language, eye contact and any other kind of communication. This allowed for the long-term accumulation of trust and understanding that allows for the reduction and correction of a wide range of communication errors. But as the groups involved grew, and the time devoted to each interaction therefore shortened, fewer channels of communication were available, and spoken language won out as the main channel of communication between people meeting each other infrequently and for short periods of time, mainly because spoken language is a relatively precise way to communicate concepts. Ultimately, as networks of communication grew even further, the need to avoid misunderstandings and errors must also have had an impact on language itself, requiring the communities concerned to develop more and more precise ways of expressing themselves in a shorter and shorter time. That impact, it seems to me, must have been visible in a proliferation of more and more, but ever narrower, concepts (categories) at any particular level of abstraction—thus reducing the number of dimensions in which these concepts could be interpreted. The multiplicity of meanings attached in different contexts to the same words—or the same roots—that one sees in any etymological dictionary bears testimony to this process, as does the proliferation of artifact categories through time, each with more and more precise and limited functions. Simultaneously, an increase in the number of levels of abstraction itself did compensate for this fragmentation, so that one could still find ways to 'lump' over these increasingly narrow concepts along crosscutting dimensions. 'Information' is but one of the last major abstractions introduced.

In western science, a similar process of fragmentation has been observable at least since the 14th century, and for very similar reasons (*cf.* Evernden, 1992). During these centuries, science has emphasized the need to solidify as much as possible the relationship between observations and interpretations, and thus between the realm of the real, with its infinite number of dimensions, and the realm of ideas, in which only a limited number of dimensions is cognized. Much scientific explanation therefore consisted of reducing the large number of dimensions involved in the processes observed into a much more limited number that was manageable in the (individual or collective) human brain, and could thus be shaped into a coherent and comprehensible narrative. Hence the fact that such science was generally '*reductionist*'. A corollary of this is the fact that, particularly in empirical science, each complex phenomenon was 'broken up' into component parts in the hope that once these components had been explained, they could be put together to explain the whole phenomenon in all its complexity. This led to the same kind of *fragmentation* that occurred in languages in general, observable at the highest level in the current division of human inquiry into disciplines, sub-disciplines, specializations, etc., each practiced by its own community that has developed its own epistemology, perspective, language, concepts, methods, techniques and values.

We now see that fragmentation as one of the main handicaps in our attempts to understand the full complexity of the processes going on around us. Moreover, the interpretations linked the phenomena investigated to processes that preceded the time at which these phenomena were observed, rather than to what was still to come (and therefore could not be observed). Scientific reasoning therefore emphasized the explanation of extant phenomena in terms of *chains of cause-and-effect* and (much later) an emphasis on feed*back* loops, in both cases linking the progress of processes through time to their antecedent trajectory. In particular, it has emphasized thinking about "origins" rather than "emergence", about "feedback" rather than "feed-forward", about "learning from the past" rather than "anticipating the future". Hence, it is no surprise that 'thinking about the future', whether one calls it 'futurology', 'forecasting', 'scenario construction' or 'foresighting' is actually a stepchild in our current academic and research institutions, and is principally developed in industry or government.

As a result of these tendencies, both in our societies' communication and culture, and in our scientific research, we have now come to a point where the unanticipated consequences of our interventions in the environment threaten to overwhelm us because of their complexity. So many unknown dimensions are involved in the dynamics of our socio-natural environment that we increasingly feel we no longer have any means to understand, limit or control their effects. That feeling is experienced as a 'crisis', and we encounter it more and more frequently—whether in the financial domain, or in those of food security, natural hazards, the security of our societies from terrorism or other undermining activities, etc.

One could effectively define such 'crises' as temporary incapacities of our society to process the information necessary to deal adequately with the external and internal dynamics it is engaged in. In our perspective, these incapacities are the result of the fact that the gap between the number of dimensions cognized in the society, and the number of dimensions playing a role in the socio-natural dynamics it is involved in, is crossing a threshold beyond which the former is inadequate to deal adequately with the latter. In the run-up to that threshold, a clear 'early warning' signal is the fact that society increasingly suffers from 'short-termism', a focus on the immediate challenges that it encounters, without taking the longer term into account: in other words, the fact that tactics come to prevail over strategy in much decision-making.

The core of the challenge seems to be that we must find ways to turn lessons *from* the past into lessons *for* the future! To do so, we must devise ways to argue coherently—and as far as possible falsifiably in Popper's (1959) sense—from the simple to the complex in order to better anticipate the complex consequences of our actions. That would enable us to re-emphasize long term, strategic thinking and a holistic vision that favors intellectual fusion between different scientific communities and perspectives. To do so, we must crucially acquire the capacity to increase, rather than reduce, the number of dimensions that we can harness in order to understand complex phenomena, so that we may attain a better understanding of the consequences of our actions because we can consider more dimensions in our decision-making about interventions in the environment.

CONCLUSION: IS THERE A WAY OUT?

It initially seems as if our intellectual and scientific tradition, the size of our interactive population, the nature of many of our languages, the under-determination of our theories by our observations (*cf.* Atlan, 1992; van der Leeuw) and the limitations of our human short-term working memory are as many challenges to our capacity to fundamentally change the nature of our thinking, and more specifically to our capacity to explicitly focus on the future and extrapolate new dimensions from the ones we know at any particular point in time. There are many examples of individuals or (small) groups of people who have nevertheless done so with some degree of success, from classical Greek philosophers via Leonardo da Vinci to 18th and 19th century science-fiction authors (such as Jules Verne or Paul Deleutre[3]). They have been able to design utopias or to extrapolate positively from their lifetime observations into the future, even though some of these ideas were never implemented or only realized years or centuries later. Inventors have also been able to anticipate, and most of us call on our "intuition" when we need to do so.

Moreover, there are some (shy) beginnings of a wider trend in this direction that we can point to. The kind of reductionist, fragmented and 'explanatory' science that resulted from these tendencies has in the past twenty-five years come under increasing attack from the 'Complex Systems' perspective emerging in the 1980's (e.g. Mitchell, 2009). It assumes that in order to get a realistic representation of reality, we need to study emergence, feed-forward and develop a generative perspective to which the amplification of the number of cognized dimensions is essential. In other quarters, 'foresighting' is spreading from the relatively limited field of industrial and economic decision-support tools to academic practitioners who actually delve into the epistemological and other challenges that

[3] Writing under the pseudonym Paul d'Ivoi, this French author anticipated the idea of modern telecommunications (wireless and television)

need to be met for this kind of science to flourish (Wilkinson and Eidinow, 2008; Selin, 2006). And yet elsewhere, under pressure from the looming environmental challenges of the 21st century, the scientific community is beginning to look ahead at 'unanticipated consequences' and what these may imply for the challenges of the future (*e.g.* Ostrom, 2009). This seems to indicate that the current predicament is more due to over-investment in the long-standing reductionist approach than anything more fundamental, and that, at least in theory, it should be possible to transcend our relative incapacity to deal with the complexities of the dynamics we are involved in.

Overcoming the limitations of human STWM

Although I am not an expert in the field at all, it seems to me that the ICT revolution has indeed created the conditions for us to overcome the limitations to our cognitive capacities that are inherent in our short-term working memory. Present-day computers do have the capacity to deal with an almost unlimited number of dimensions and information sources in real time, and thus to overcome what appeared at first sight to be the most fundamental of the barriers mentioned above. But that capacity has not been fully exploited because of our long-standing and ubiquitous scientific and intellectual tradition, which has emphasized the use of such equipment as part of the process of dimension-reduction that provides acceptable explanations, rather than as a tool to increase the number of dimensions taken into account in our understanding of complex phenomena. Under the impact of complex systems science this is clearly changing (as seen, for example, in the increased use of high-dimensional Agent Based Models, but much more needs to be done, mainly in developing conceptual and mathematical tools as well as appropriate software.

Overcoming the under-determination of our theories by observations

Similarly, and with the same caveat that I am not a professional in this field, I am under the impression that the very recent revolution in IT capacity to continuously monitor processes on-line, and to treat and store the exponentially increased data streams that are generated by such monitoring, points to the fact that we may indeed be on the brink of (at least partly) overcoming the under-determination of our theories by our observations, and that this is the corollary of the dimension-reduction traditional science practices (Atlan, 1992). The reduction in the size and cost of the monitoring equipment is quickly bringing such massive data collection within reach. Simultaneously, the development of novel data-mining techniques is helping us to make sense of the data thus collected, or at least in selecting the appropriate data to be scrutinized in order to better inform our theories.

Transforming our scientific and intellectual tradition

Although I am not among those who fall easily for panaceas, I do believe that the complex (adaptive) systems approach is a useful first step on the way to fundamentally transforming our scientific and intellectual tradition from studying stasis and preferring simple to complex explanations, to studying dynamics, with an emphasis on emergence and inversion of Occam's razor (increasing the number of dimensions taken into account). Clearly, we have a long way to go in this domain, but the rapid and substantive advances in certain fields, including physics, biology and economics, coupled with the rapid recent spread of this approach in universities in many parts of the world and the growing awareness of the need for more holistic approaches in such domains as sustainability and health, cause me to be moderately optimistic about our chances of transforming our scientific and intellectual tradition.

The communication challenge

The underlying communication challenge is how to communicate other than linearly and in writing or speech with an increasingly large number of partners at very variable distances. This is the trend that was, in my opinion, responsible for the particular development referred to above: narrower and narrower concepts, and the consequent fragmentation of our perspective on the world. Contrary to some, I do not think language is subject to deliberate change—it adapts itself to human needs and ideas in a 'bottom-up' process. But even if it were possible to transform the ways in which we speak and write, we would still have an essentially linear communication tool. The question is therefore whether the radically different ways of interactively communicating that are made possible by modern communications technologies, and in particular the collective building of knowledge using multimedia, as is made possible in web 2.0, will allow us to communicate non-linearly and in more dimensions. This would entail the directed use of visuals, which generally can communicate more dimensions simultaneously than spoken or written language.

Transforming our thinking

The kind of reductionist thinking that I am referring to is so heavily ingrained and so widely spread in our culture and our kinds of science that changing our thinking will require a major effort. Our world view, our language, our institutions all militate against such a change, and most importantly, we are for the moment lacking a coherent alternative way of thinking against which we can leverage our present-day science. By far the greatest challenge from the perspective of human and financial capital and effort therefore appears to me to be in the domain of education, from the earliest childhood throughout university and into adult life. The current education system in the developed world is, overall, no longer adapted to the challenges of the 21st century, among which sustainability looms so large. We have to move away from knowledge acquisition aimed at question-driven research towards challenge-focused education that aims to help deal with substantive challenges, from 'linear explanation' in terms of cause-and-effect to 'multi-dimensional projection' in terms of alternatives, from one-to-many teaching (in which an instructor tells students what to do, what is right and what is wrong), to many-to-many teaching in which instructors and students all interact, learn and teach. At the same time, we must develop education systems that stimulate the acquisition of creativity, risk-taking and diversity rather than conformity and risk-adverseness. In doing so we must harness the tools referred to above, but more than anything we must 'bend' minds around to thinking in new, uncharted, ways. In doing so, we are handicapped by the fact that economics, career structures, evaluations, disciplinary momentum and many other factors and dynamics are stacked against success in this area. There is a lot of work to be done!

REFERENCES

ALP, I. E. (1994), "Measuring the size of working memory in very young children: the imitation sorting task", *International Journal of Behavioral Development* 17, pp. 125–141.

ATLAN, H. (1992), "Self-organizing networks: weak, strong and intentional. The role of their under-determination", *La Nuova Critica*, N. S., 19-20(1/2), pp. 51-70.

BATESON, G. (1972), *Steps to an Ecology of Mind*, New York: Ballantine

BETTENCOURT, L. M. A., J. LOBO, D. HELBING, C. KÜHNERT and G.B. WEST (2007), "Growth, Innovation, Scaling and the Pace of Life in Cities", *Proceedings of the National Academy of Sciences* (US) 104(17), pp. 7301-7306.

BOYD, R., and P. J. RICHERSON (1985), *Culture and the Evolutionary Process,* Chicago: University of Chicago Press

CARLSON, S. M., L. J. MOSES and C. BRETON (2002), "How Specific is the Relation between Executive Function and Theory of Mind? Contributions of Inhibitory Control and Working Memory", *Infant-Child Development* 1, pp. 73–92.

CORRÊA, L. M. S. (1995), "An alternative assessment of children's comprehension of relative clauses", *Journal of Psycholinguistic Research* 24, pp. 183–203.

CRONON, W. (1983), *Changes in the Land: Indians, Colonists, and the Ecology of New England*, New York: Hill and Wang.

DIAMOND, A. and B. DOAR (1989), "The Performance of Human Infants on a Measure of Frontal Cortex Function, the Delayed-Response Task", *Developmental Psychobiology* 22, pp. 271–294.

EPSTEIN, H. T. (2002), "Evolution of the Reasoning Brain", *Behavioral Brain Science* 25, pp. 408–409.

EVERNDEN, N. (1992), *The Social Creation of Nature*, Baltimore: Johns Hopkins Press.

JOHNSON, J., V. FABIAN and J. PASCUAL-LEONE (1989), "Quantitative Hardware Stages that Constrain Language Development", *Human Development,* 32, pp. 245–271.

KEMPS, E., S. De Rammelaere and T. Desmet (2000), "The Development of Working Memory: Exploring the Complementarity of Two Models", *Journal of Experimental Psychology* 77, pp. 89–109.

KIDD, E. and E. L. BAVIN (2002), English-speaking Children's Comprehension of Relative Clauses: Evidence for General–Cognitive and Language-Specific Constraints on Development", *Journal of Psycholinguistic Research,* 31, pp. 599–617.

LANE, D., R. MAXFIELD, D. W. READ and S. E. VAN DER LEEUW (2009), "From Population Thinking to Organization Thinking", in D. Lane, D. Pumain, S. E. van der Leeuw and G. West (eds.), *Complexity Perspectives on Innovation and Social Change*, pp. 11-42, Berlin: Springer (Methodos series).

LUCIANA, M. and C. A. NELSON (1998), "The Functional Emergence of Prefrontally-Guided Working Memory Systems in Four-to-Eight-Year-Old Children", *Neuropsychologia* 36, pp. 273–293.

MARTIN, R. D. (1981), "Relative Brain Size and Basal Metabolic Rate in Terrestrial Vertebrates", *Nature* 293, pp. 7–60.

MITCHELL, M. (2009), *Complexity: a Guided Tour,* New York: Oxford University Press

OSTROM, E. (2009), "A General Framework for Analyzing Sustainability of Social-Ecological Systems", *Science* 325(5939), pp. 419-422.

PIGEOT, N. (1991), "Reflexions sur l'histoire technique de l'homme: De l'évolution cognitive à l'évolution culturelle", *Paléo* 3, pp. 167-200.

POPPER, K. (1959), *The Logic of Scientific Discovery*, London: Hutchinson.

READ, D. W., D. A. LANE and S. E. VAN DER LEEUW (2009), "The Innovation Innovation", in (D. A. Lane, D. Pumain, S. E. van der Leeuw and G. West (eds.) *Complexity Perspectives on Innovation and Social Change* Berlin: Springer Verlag.

READ, D. W. and S. E. VAN DER LEEUW (2008), "Biology Is Only Part Of The Story…", *Philosophical Transactions of the Royal Society, Series B* 363, pp. 1959-1968.

READ, D. W. and S. E. VAN DER LEEUW (2009), "Biology Is Only Part Of The Story…", in A. C. Renfrew and L. Malafouris (eds.), *The Sapient Mind*, Oxford: Oxford University Press, pp. 33-49.

RIGHTMIRE, G. P. (2004), "Brain Size and Encephalization in Early to Mid-Pleistocene Homo", *American Journal of Physical Anthropology* 124, pp. 109–123.

RUFF, C. B., E. TRINKHAUS and T. W. HOLLIDAY (1997), "Body Mass and Encephalization in Pleistocene Homo", *Nature* 387, pp. 173–176.

SELIN, C. (2006), "Trust and the Illusive Force of Scenarios", *Futures* 38(1), pp. 1-14.

SHANNON, C. E. and W. WEAVER (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal* 27, pp. 379-423 and 623-656.

SIEGEL, L. S. and E. B. RYAN (1989), "The Development of Working Memory in Normally Achieving and Subtypes of Learning-Disabled Children", *Child Development* 60, pp. 973–980.

TAINTER, J. A. (1988), *The Collapse of Ancient Societies,* Cambridge (UK); Cambridge University Press.

VAN DER LEEUW, S. E. (1986), "On Settling Down and Becoming a 'Big-Man'", in M. A. van Bakel, R. R. Hagesteijn and P. van de Velde (eds.), *Private Politics: a Multi-Disciplinary Approach to 'Big-Man' Systems*, Leiden: Brill Academics Publishers, pp. 33-47.

VAN DER LEEUW, S. E. (1990), "Archaeology, Material Culture and Innovation", *SubStance* 62-63, pp. 92-109.

VAN DER LEEUW, S. E. (2000), "Making Tools from Stone and Clay", in A. Anderson and T. Murray (eds.), *Australian Archaeologist: Collected Papers in Honour of Jim Allen*, Coombs: Academic Publishing.

VAN DER LEEUW, S. E. (2007), "Information Processing and its Role in the Rise of the European World System", in R. Costanza, L. J. Graumlich and W. Steffen (eds.), *Sustainability or Collapse?*, Cambridge, MA: The MIT Press, Dahlem Workshop Reports, pp. 213-241.

WILKINSON, A. and E. EIDINOW (2008), "Evolving Practices in Environmental Scenarios: a New Scenario typology", *Environmental Research Letters,* 3(4) 045017.

Innovation: It is generally agreed that science shapes technology, but is that the whole story?

Nathan Rosenberg
Stanford University
Stanford Institute for Economic Policy Research

Not to prolong your suspense, the correct answer to the question in the subtitle of my paper is the obvious one: causation runs both ways. But I want to persuade you that the causation running from technology to science is vastly more powerful than is generally realized.

The reasoning is straightforward. A market economy generates powerful incentives to undertake certain kinds of scientific research. This is because the eventual findings of such research can be made to improve the performance, or to reduce the cost, of technologies that are vital to the competitive success of profit-making firms. Further, I want to suggest that there were powerful forces at work in the course of the 20th century that had the effect of expanding the ways in which changes in the realm of technology have led to changes in the various realms of science. I want to call your attention to some of the most significant organizational changes, and associated changes in incentives, that were responsible for strengthening the causal forces that flowed from technology to science.

In order to do this, I will need to introduce just one single bit of jargon: I will use the term "endogenous" from the perspective of the economist and not from the perspective of the scientist. Thus, when I refer to the endogeneity of science, I am referring to the extent to which scientific progress has been directly influenced by the working out of the normal forces of the market place. My justification is that I will be trying to identify forces that emerged in the course of the twentieth century that made scientific research more responsive to economic incentives.

I also need to emphasize one caveat that I cannot emphasize too strongly. I am not implicitly suggesting that the financial support of the country's scientific research should be left to the market place. Rather, I will be calling attention to the operation of market forces that have become increasingly supportive of scientific research. I believe that these developments were crucial to the rapid expansion of American industry, but that is very different from suggesting that market forces, by themselves, were sufficient.

CORPORATE RESEARCH LABS

The proposition that scientific research became increasingly endogenous in the

> "A market economy generates powerful incentives to undertake certain kinds of scientific research. This is because the eventual findings of such research can be made to improve the performance, or to reduce the cost, of technologies that are vital to the competitive success of profit-making firms"

course of the 20th century must necessarily begin by focusing on a key organizational innovation: the industrial research lab. It was these corporate labs that determined the extent to which the activities of the scientific community could be made responsive to the needs of the larger economy. But such a statement, by itself, cannot stand alone. That is because these research labs depended for their effective performance, in turn, upon a network of other institutions. These included, above all, research at universities. Before the Second World War, university research depended heavily for its financing on private philanthropic foundations, such as the Rockefeller, Guggenheim and Carnegie foundations. In the pre-war period, as well, universities often relied on financial support from local industry for carrying out certain certain classes of research, mostly of an applied nature. This was especially true of state universities, where it was essential to provide evidence of assistance to local industry [agriculture, mining, railroads] in order to justify the imposition of taxes upon the citizens of each state. In fact, with few exceptions, funds raised by state governments went overwhelmingly to support teaching and not research.

This situation was totally transformed in the post World War II period when the federal government became, overwhelmingly, the dominant patron of scientific research, and universities became the primary locus of such research. It is important to note that the concentration of basic scientific research in the university community where, I think it is fair to say, it has flourished, has been an organizational arrangement almost unique to the US. Unlike the situation in Western Europe, where basic research has been concentrated in government labs (Max Planck, CNRS), federal laboratories in US have accounted for less than 10% of basic research (9.1% in the mid-1990s).

A further distinctive feature of great importance in the US is the very large commitment of private industry to scientific research that the NSF defines as basic. Private industry accounted for slightly over 30% of all basic research in the year 2000 (probably declining slightly in the last few years). Although at last count there were around 16,000 private firms that had their own corporate labs, the vast majority of these firms conduct research of a predominantly applied nature. Only a very small number do basic research. Nevertheless, over the years, a few of these corporate labs have conducted research of the most fundamental nature—General Electric, IBM and, most important of all, Bell Labs before the divestiture of AT&T in 1984. Researchers in a number of corporate labs have won Nobel Prizes: most recently Jack Kilby, of Texas Instruments, won the Prize for Physics in the year 2000, for research leading to the development of the integrated circuit [Kilby's research received financial support from the federal government].

Having said this, it is essential to realize that the research activities of industrial labs should not be evaluated, as they often are by academics, by the usual academic criteria—such as publications in prestigious professional journals or the winning of Nobel Prizes. Such labs have a very different purpose. The industrial lab is essentially an institutional innovation (of German origin) in which the research agenda is largely shaped by the short-term needs but also, in a few notable cases, by the longer-term strategies of industrial firms. Within the industrial context, the intended role of corporate scientists is to improve the performance of their respective firms in the competitive context of (mostly) high tech sectors of the economy. Thus the critical achievement of the growth of the American industrial lab in the course of the 20th century has been to subject science, more and more, to commercial criteria. In so doing, it rendered science an activity whose directions were increasingly shaped by economic forces and were concentrated on the achievement of economic goals—which is to say that such scientific research should be regarded as largely endogenous.

One further strategic role of the corporate lab arises from the fact that a firm cannot effectively monitor and evaluate the findings, and the possible implications, of the huge volume of university research unless it has its own internal capability for doing such things. The importance of this point cannot be overestimated. In advanced industrial societies that are now simply flooded with the flow of information, not only from universities, but from professional journals on library shelves or electronically via Internet search engines such as Yahoo and Google, the exploitation of this vast flow of information requires an internal competence that, typically, only in-house scientists can provide. Indeed, America's remarkable commercial successes in high tech markets over the past 50 years have owed a great deal to these internal competences in private industry. Industrial scientists have played a critical role in the transfer of potentially useful knowledge generated by university research, not only because of their scientific sophistication, but also because they have had a deep awareness of their firms' commercial priorities and technological capabilities (Rosenberg, 2002; Mowery and Rosenberg, 1998).

HOW ENGINEERING DISCIPLINES HAVE SHAPED SCIENCE

I would now like to call your attention to another major force for advancing the endogeneity of science in the course of the 20th century. I would like to pose the question: what specific role is played by engineering disciplines in determining the scientific agenda of private firms? Let me respond first by offering a clarification. It is a common practice to characterize engineering disciplines as being essentially applied science. This is, in my view, a seriously misleading characterization. A more careful unwinding of the intertwining of science and technology suggests that the willingness of profit-seeking firms to devote money to scientific research is very much influenced by the prospect of converting such research findings into finished and marketable products. The actual conduct of scientific research may not be undertaken with highly specific objectives in mind, but rather with an increased confidence that, whatever the specific research findings, an enlarged engineering capability will substantially increase the likelihood of being able to use these findings to bring improved or new products to the market place.

From this perspective, there is a serious sense in which the economist may argue that the science of chemistry should be thought of as an application of chemical engineering! Alternatively put, the growing sophistication of engineering disciplines has had the result of strengthening the endogeneity of science. I do not want this point to be made to sound too paradoxical. I mean to suggest that the willingness of private industry to commit financial resources to long-term scientific research has been considerably strengthened by the progress of the appropriate engineering disciplines. Such progress raises the confidence of corporate decision makers that the findings of basic research may eventually be converted to profitable uses.

This argument seems particularly pertinent to the specialty of polymer chemistry, a field that was opened by the researches of Staudinger, Meyer and Mark in Germany in the 1920s. In the US at least, polymer chemistry is a field that has long been dominated by the industrial research community. The fundamental research contributions to polymer chemistry of Wallace Carothers at du Pont, beginning in 1928, owed a great deal to the increasing maturity of chemical engineering in the preceding decade or so—an engineering discipline to which du Pont had made important contributions (Hounshell and Smith). Carothers' research findings led directly to the discovery of nylon, the first of a proliferation of synthetic fibers that came to constitute an entirely new subsector of the petrochemical industry after the Second World War. But it is doubtful whether du Pont would have committed itself to Carothers' costly, fundamental researches in polymer chemistry in the first place, in the absence of the progress in chemical engineering in the decade preceding 1928. Thus, progress at the technological level (chemical engineering) increasingly strengthened the willingness to spend money on science, which I regard as a growth in the endogeneity of science (Rosenberg, 1998).

Let me sketch out the intermediate steps that underlie my argument. The discipline of chemical engineering really had its beginnings in the second and third decades of the 20th century, mainly at MIT in response to the spectacular expansion of the automobile industry and, along with that industry's growth, a voracious demand for refined chemical products (primarily, of course, for high octane gasoline). The scale of that growth can be captured in the following numbers: In 1900 the automobile industry was so insignificant that the Census Bureau classified cars under the category "Miscellaneous." [In that year there were only 8,000 registered cars in the US]. By 1925 the automobile industry had leaped to the status of the largest manufacturing industry in the whole country (measured by value added).

It was the growth of the automobile that gave birth to the discipline of chemical engineering. Chemical engineers, during the 1920s and later, transformed the petroleum refining industry from small-scale batch production into one of vastly larger scale and continuous processing. The emerging chemical engineering discipline accomplished this by developing a new conceptual framework within which it became possible to introduce scientific concepts and methodologies from such fields as fluid flow (fluid dynamics), heat transfer and, in the 1930s, the pervasive power of thermodynamics. In other words, the design of chemical process plants could now draw heavily upon a number of different scientific realms. Thus, it was the establishment of a new engineering discipline, in responding to the rapid expansion of a new transportation technology that, in turn, laid the basis for the profitability of scientific research, not only in du Pont and petroleum-refining firms, but in a very wide range of industries that also made use of chemical process plants. It is worth emphasizing how pervasive chemical process plants became in the course of the 20th century. Large chemical plants could be found in petroleum refining, rubber, leather, coal (by-product distillation plants), food-processing, sugar refining, explosives, ceramics and glass, paper and pulp, cement, and metallurgical industries (e.g., aluminum, iron and steel).

> "The actual conduct of scientific research may not be undertaken with highly specific objectives in mind, but rather with an increased confidence that, whatever the specific research findings, an enlarged engineering capability will substantially increase the likelihood of being able to use these findings to bring improved or new products to the market place"

HOW NEW PRODUCTS HAVE SHAPED SCIENCE

The next related observation with respect to the growing endogeneity of scientific research goes beyond the role played by engineering disciplines in strengthening the private incentives to perform scientific research. The argument here is that the development of some specific new product, that is perceived to have great commercial potential, may provide, and often has provided, a powerful stimulus to scientific research. This proposition is surprising only if one is already committed to a rigid, overly simplistic linear view of the innovation process—one in which causality is always expected to run from prior scientific research to "downstream" product design and engineering development. There is in fact, however, a straightforward endogenous explanation at work here. A major technological breakthrough typically provides a strong signal that a new set of profitable opportunities has been opened up in some precisely-identified location. Consequently, it is understood that scientific research that can lead to further improvements in that new technology may turn out to be highly profitable.

The problems encountered by sophisticated industrial technologies, and the anomalous observations and unexpected difficulties that they have encountered, have served as powerful stimuli to much fruitful scientific research in the academic community as well as the industrial research laboratory. In these ways the responsiveness of scientific research to economic needs and technological opportunities has been powerfully reinforced.

This was dramatically demonstrated in the case of the advent of the transistor, the discovery of which was announced at Bell Labs in the summer of 1948. Within a decade of that event solid-state physics, which had previously attracted the attention of only a small number of researchers and was not even taught at the vast majority of American universities (mainly MIT, Princeton, and Cal Tech) had been transformed into the largest sub-discipline of physics. It was the development of the transistor that changed that situation by dramatically upgrading the potential financial payoff to research in the solid state. J. A. Morton, who headed the fundamental development group that was formed at Bell Labs after the invention of the transistor, reported that it was extremely difficult to hire people with a knowledge of solid-state physics in the late 1940s. Moreover, it is important to emphasize that the rapid mobilization of intellectual resources to perform research in the solid state occurred in the university community as well as in private industry, immediately after the announcement of the momentous findings of Shockley and his research colleagues at Bell Labs. As one strong piece of evidence for this view, the number of publications in semiconductor physics rose from less than 25 per annum before 1948 to over 600 per annum by the mid-1950s (Herring).

The chronology of the events to which I have just referred is essential to my argument. Transistor technology was not the eventual consequence of a huge prior buildup of resources devoted to solid-state physics, although it was of course also true that some of the twentieth century's most creative physicists had been devoting their considerable energies to the subject. Rather, it was the initial breakthrough of the transistor, as a functioning piece of hardware, that set into motion a vast subsequent commitment of financial support for scientific research. Thus, the difficulties that Shockley encountered with the operation of the early point-contact transistors led him into a systematic search for a deeper explanation of their behavior, expressed in terms of the underlying quantum physics of semiconductors. This search not only led eventually to a vastly superior amplifying device, the junction transistor; it also contributed to a much more profound understanding of the science of semiconductors. Indeed, Shockley's famous and highly influential book, *Electrons and Holes in Semiconductors*, drew heavily upon this research, and the book was the direct outgrowth of an in-house course that Shockley had taught for Bell Labs' personnel. Moreover, Shockley also found it necessary to run a six day course at Bell Labs in June 1952 for professors from some thirty universities as part of his attempt to encourage the establishment of university courses in transistor physics.

Clearly, the main flow of scientific knowledge during this critical period was from industry to university, and not the other way around. Indeed, for a considerable period of time, Stanford and the University of California at Berkeley had to employ scientists from local industry to teach courses in solid-state physics/electronics.

A similar sequence can be seen in the commitment of funds to research in surface chemistry, after problems with the reliability

> “Under modern industrial conditions, technology has come to shape science in the most powerful of ways: by playing a major role in determining the research agenda of science as well as the volume of resources devoted to specific research fields”

of early transistors pointed in that direction. More recently, and to compress a much more complex chain of events, the development of laser technology suggested the feasibility of using optical fibers for telephone transmission purposes. This possibility naturally pointed to the field of optics, where advances in scientific knowledge could now be expected to have potentially high economic payoffs. As a result, optics as a field of scientific research experienced a great resurgence in the 1960s and after. It was converted by changed expectations, based upon recent and prospective technological innovations, from a relatively quiet intellectual backwater of science into a burgeoning field of research. This growth of activity in the discipline was generated not by forces internal to the field of optics, but by a radically altered assessment of the potential opportunities for laser-based technologies. Moreover, different kinds of lasers gave rise to different categories of fundamental research. As Harvey Brooks noted, “While the solid-state laser gave a new lease of life to the study of insulators and of the optical properties of solids, the gas laser resuscitated the moribund subject of atomic spectroscopy and gas-discharge physics” (Brooks, 1968).

I draw the conclusion from this examination that, under modern industrial conditions, technology has come to shape science in the most powerful of ways: by playing a major role in determining the research agenda of science as well as the volume of resources devoted to specific research fields. One could examine these relationships in much finer detail by showing how, throughout the high tech sectors of the economy, shifts in the technological needs of industry have brought with them associated shifts in emphasis in scientific research. When, for example, the semiconductor industry moved from reliance upon discrete circuits (transistors) to integrated circuits, there was also a shift from mechanical to chemical methods of fabrication. When Fairchild Semiconductors began to fabricate integrated circuits, they did so by employing new methods of chemical etching that printed the transistors on the silicon wafers and also laid down the tracks between them. This chemical technique did away with expensive wiring, and also produced integrated circuits that operated at much higher speeds. At the same time, the increased reliance upon chemical methods brought with it an increased attention to the relevant subfields of chemistry, such as surface chemistry.

I cite the experience of changing methods of wafer design and fabrication to indicate the ways in which the changing needs and priorities of industry have provided the basis for new priorities in the world of scientific research. But it is essential to emphasize that these new priorities exercised their influence, not only upon the world of industrial research, but upon the conduct of research within the university community as well. I need only point out that Stanford University has for some time had, its own Center for Integrated Systems. This Center is devoted to laboratory research on microelectronic materials, devices, and systems, and is jointly financed by the federal government and private industry.

SERENDIPITY

There is a further source of causation running from technology to science to which I would like to call your attention. I refer to the role of serendipity. It is, of course, to be expected that well-trained scientific minds are likely to turn up unexpected findings in many places. As Pasteur expressed it in the mid-19th century, "Where observation is concerned, chance favors only the prepared mind." By way of contrast, consider Thomas Edison, by universal consent a brilliant inventor, but someone who had little interest in observations that had no immediate practical relevance. In 1883 he observed the flow of electricity across a gap, inside a vacuum, from a hot filament to a metal wire. Since he saw no practical application and had no scientific training, he merely described the phenomenon in his notebook and went on to other matters of greater potential utility in his effort to enhance the performance of the electric light bulb. Edison was, of course, observing a flow of electrons, and the observation has since even come to be referred to as the "Edison Effect"—named after the man who, strangely enough, had failed to discover it. Had he been a curious (and patient) scientist, less preoccupied with matters of short-run utility, Edison might later have shared a Nobel Prize with Owen Richardson who analyzed the behavior of electrons when heated in a vacuum, or conceivably even with J. J. Thomson for the initial discovery of the electron itself. Edison's "prepared mind," however, was prepared only for observations that were likely to have some practical relevance in the short run.

A distinctive feature of the 20th century in dynamic capitalist economies was the vastly-increased numbers of scientifically "prepared minds" in both the universities and private industry. The pursuit of the possible implications of unexpected observations became the basis on many occasions for fundamental breakthroughs that occurred serendipitously when "prepared minds" were available to pursue the possible implications of the unexpected. Surely the most spectacular instance of serendipity in the 20th century—not achieved in an industrial laboratory—was Alexander Fleming's brilliant conjecture, in 1928, that the unexpected bactericidal effect that he had observed in the bacterial cultures in his Petri dish, was caused by a common bread mould that had accumulated on his slides. Fleming published this finding in 1929, but no substantial progress was made in producing a marketable product until more than a decade later, when the exigencies of wartime led to a joint, Anglo-American "crash" program to accelerate the production of the antibiotic (Elder, 1970).

It is at least a plausible speculation that, had Fleming made his marvelous discovery while working in a pharmaceutical lab, penicillin would have become available, in large quantities, far more swiftly than was in fact the case. For a contrary view, see Bernal: vol. 3, 926-927). In the context of this paper it is also worth pointing out a little-known historical fact, that the technology to produce the antibiotic in bulk was achieved not, as would ordinarily have been expected, by a pharmaceutical chemist, but by chemical engineers. It was the chemical engineers who demonstrated how a technique called "aerobic submerged fermentation," which became the dominant production technology, could be applied to this complex product (Elder, 1970).

The growth of organized industrial labs in 20th century America vastly enlarged the number of trained scientists in the industrial world who encountered strange phenomena that were most unlikely to occur, or to be observed, except in some highly specialized industrial context. In this sense, the huge increase in new high-tech products, along with dense concentrations of well-trained scientific specialists in industry, sharply increased the likelihood of serendipitous discoveries in the course of the twentieth century.

Consider the realm of telephone transmissions. Back at the end of the 1920s, when transatlantic radiotelephone service was first established, the service was discovered to be poor due to a great deal of interfering static. Bell Labs asked a young man, Karl Jansky, to determine the source of the noise so that it might be reduced or eliminated. He was given a rotatable antenna to work with. Jansky published a paper in 1932 in which he reported that he had found three sources of noise: Local thunderstorms, more distant thunderstorms, and a third source which he described as "a steady hiss static, the origin of which is not known". It was this "star noise" as Jansky labelled it, which marked the birth of the entirely new science of radio astronomy.

Jansky's experience underlines why the frequent attempt to distinguish between basic research and applied research is extremely difficult to make in practice. Fundamental scientific breakthroughs often occur while dealing with very applied or practical problems, especially problems relating to the performance of new technologies in an industrial context.

But the distinction breaks down in another way as well. It is essential to distinguish between the personal motives of the individual researchers and the motives of the decision makers in the firm that employs them. Many scientists in private industry could honestly say that they are attempting to advance the frontiers of basic scientific knowledge, without any concern over possible applications. At the same time, the motivation of the research managers, who decide whether or not to finance research in some basic field of science, may be strongly motivated by expectations of eventual useful findings.

This certainly appears to have been the case in the early 1960s when Bell Labs decided to support research in astrophysics because of its potential relationship to the whole range of problems and possibilities in the realm of microwave transmission and especially in the use of communication satellites for such purposes. It had become apparent that, at very high frequencies, annoying sources of interference in transmission were widely encountered.

This source of signal loss was a matter of continuing concern in Bell Labs' development of the new technology of satellite communications. It was out of such practical concerns that Bell Labs decided to employ two astrophysicists, Arno Penzias and Robert Wilson. Penzias and Wilson would undoubtedly have been indignant if anyone had suggested that they were doing anything other than basic research. They first observed the cosmic background radiation, which is now taken as confirmation of the "Big Bang" theory of the formation of the universe, while they were attempting to identify and measure the various sources of noise in their antenna and in the atmosphere. It seems fair to say that this most fundamental breakthrough in cosmology in the past century was entirely serendipitous. Although Penzias and Wilson did not know it at the time, the character of the background radiation that they discovered was just what had been postulated earlier by cosmologists at Princeton who had devised the Big Bang theory. Penzias and Wilson shared a Nobel Prize in Physics for this finding. Their findings were as basic as basic science can get, and were in no way diminished by observing that the firm that had employed them did so because the decision makers at Bell Labs hoped to improve the quality of satellite transmission.

The parallelism between the fundamental discoveries of Jansky and Penzias and Wilson is, of course, very striking. In both episodes, the Bell Labs researchers stumbled upon discoveries of the greatest possible scientific significance while involved in projects that were motivated by the desire of Bell Labs to improve the quality of telephone transmission. In the case of Penzias and Wilson, they were conducting their research with a remarkably sensitive horn antenna that had been built for the Echo and Telstar satellite projects. Wilson later stated that he was originally attracted to work at Bell Labs because working in the Labs would provide access to a horn antenna which was one of the most sensitive of such antennas in existence (Aaronson, 1979: 13).

I have called attention to two episodes at Bell Labs in which industrial researchers discovered natural phenomena of immense scientific significance, while the firm that employed them did so in the hope that they would solve serious problems connected with the performance of a new communications technology. In one sense it is fair to say that important scientific findings by profit-making firms are sometimes achieved unintentionally—they have discovered things that they were not looking for, which I take to be the generic meaning of Horace Walpole's mid-eighteenth century neologism—serendipity. Such breakthroughs in the private sector, moreover, are difficult to understand if one insists on drawing sharp distinctions between basic and applied research on the basis of the motivations of those performing the research. I find it irresistible here to invoke, once again, the shade of the great Pasteur: "There are no such things as applied sciences; only applications of science."

In fact, I would go much further: when basic research in industry is isolated from the other activities of the firm, whether organizationally or geographically, it is likely to become sterile and unproductive. Much of the history of basic research in American industry suggests that it is likely to be most effective when it is highly interactive with the work and the concerns of applied scientists and engineers within the firm. This is because the high technology industries have continually thrown up problems, difficulties and anomalous observations that were most unlikely to occur outside of specific high technology contexts.

The sheer growth in the number of trained scientists in industrial labs, along with the growth of new, highly complex, specialized products that appeared in the course of the 20th century, powerfully increased the likelihood of serendipitous findings. High-tech industries provide a unique vantage point for the conduct of basic research but, in order for scientists to exploit the potential of the industrial environment, it is necessary

"The sheer growth in the number of trained scientists in industrial labs, along with the growth of new, highly complex, specialized products that appeared in the course of the 20th century, powerfully increased the likelihood of serendipitous findings"

to create opportunities and incentives for interaction with other components of a firm. Bell Labs before divestiture (1984) is probably the best example of a place where the institutional environment was most hospitable for basic research. I do not suggest that Bell Labs was, in any respect, a representative industrial lab. Far from it. It was a regulated monopoly that could readily recoup its huge expenditures on research. But, perhaps even more important, it came to occupy a place on the industrial spectrum where, as it turned out, technological improvements required a deeper, scientific exploration of certain portions of the natural world that had not been previously studied.

INSTRUMENTATION

Of course my examination of the endogeneity of science has been no more than a very modest and partial sketch. Entire categories of the influence of technology upon science have been completely ignored here, such as the pervasive impact of new instrumentation, i.e., technologies of observation, experimentation and measurement. Indeed, scientific instruments may be usefully regarded as the capital goods of the research industry. Much of this instrumentation, in turn, has had its origins in the university world and, to underline the extent of the intertwining of technology and science in recent years, some of the most powerful of those instruments, such as Nuclear Magnetic Resonance, had their origins in fundamental research that was originally undertaken in order to acquire some highly specific pieces of knowledge, such as a deeper understanding of the magnetic properties of atomic nuclei. Indeed, Felix Bloch was awarded Stanford's first Nobel Prize in physics for precisely such research (Rosenberg, 1997; in the same volume, Kruytbosch, 1997). Nuclear Magnetic Resonance spectroscopy, in turn, became an invaluable tool in chemistry for determining the structure of certain molecules—e.g., hydrogen, deuterium, boron and nitrogen atoms (Kruytbosch, 1997: 32-34).

Clearly, instrumentation and techniques have moved from one scientific discipline to another in ways that have been highly consequential for the progress of science. In fact, it can be argued that a serious understanding of the progress of individual disciplines is generally unattainable in the absence of an examination of how different areas of science have influenced one another. This understanding is frequently tied directly to the development, the timing and the mode of transfer of scientific instruments among disciplines. The flow of "exports" appears to have been particularly heavy from physics to chemistry, as well as from both physics and chemistry to biology, to clinical medicine and, ultimately, to the delivery of health care. There has also been a less substantial flow from chemistry to physics and, in recent years, from applied physics and electrical engineering to health care. NMR eventually became the basis for one of the most powerful diagnostic tools of twentieth (and twenty-first) century medicine (MRI).

The transistor revolution was a direct outgrowth of the expansion of solid-state physics, but the successful completion of that revolution was in turn heavily dependent upon further developments in chemistry and metallurgy which provided materials of a sufficiently high degree of purity and crystallinity. Finally, physics has spawned sub-specialties that are inherently interdisciplinary: for example, biophysics, astrophysics and materials science.

One further point, however, is implicit in what has already been said. The availability of new or improved instrumentation or experimental technique in one academic discipline has often been the source of interdisciplinary collaboration. In some critical cases, it has involved the migration of highly trained scientists from one field to another, such as those physicists from the Cavendish Laboratory at Cambridge University who played a decisive role in the emergence of molecular biology. This emergence had depended heavily upon scientists trained in physicists' skills at Cavendish, who transferred the indispensable tool of X-ray crystallography into the very different realm of biology. Molecular biology was the product of interdisciplinary research in the special sense that scientists trained in one discipline crossed traditional scientific boundary lines and brought the intellectual tools, concepts and experimental methods into the service of an entirely new field (Judson, 1979).

The German physicist, Max von Laue, discovered the phenomenon of X-ray diffraction in 1912. Its applications in the early years were employed by William Bragg and his son, Lawrence Bragg, primarily in the new field of solid-state physics but also, later on, in developing the field of molecular biology. The main center of the methodology of X-ray diffraction was for many years the Cavendish Laboratory, presided over by Lawrence Bragg. Numerous scientists went there in order to learn how to exploit the technique, including Max Perutz (at the time a chemist), James Watson, Francis Crick, John Kendrew, all later to receive Nobel Prizes in Physiology and Medicine. The transfer of skills in X-ray diffraction was facilitated by the unusual step of the establishment of a Medical Research Council unit at the Cavendish, headed by Perutz but under the general direction of the physicist Lawrence Bragg (Crick, 1988: 23). James Watson (1968: 220) later reported Bragg's obvious delight over "...the fact that the X-ray method he had developed forty years before was at the heart of a profound insight into the nature of life itself". To infer the three-dimensional structure of very large-molecule proteins by the new technique of X-ray crystallography, which offered only two-dimensional photographs of highly complex molecules, appears to have been a hellishly difficult enterprise, but it provided much of the basis for the new discipline of molecular biology. Rosalind Franklin who, sadly, died very young, is widely agreed to have been the most skilful practitioner of X-ray crystallography.

Moreover, it is important to observe that the two separate communities—university scientists (including medical school clinicians) and commercial instrument makers—interacted with and influenced one another in ways that were truly symbiotic.

Precisely because these two communities marched to the tunes of very different drummers, each was ultimately responsible for innovative improvements that could not have been achieved by the other, had the other been acting alone (Gelijns and Rosenberg, 2000). It should be added that the applications of physics research have usually moved more readily across disciplinary boundary lines in industry than they have in the academic world. Profit-making firms are not particularly concerned with where those boundary lines have been drawn in the academic world; they tend to search for solutions to problems regardless of where those solutions might be found (NRC, 1986).

Thus, the technological realm has not only played a major role in setting the research agenda for science, as I have argued. Technology has also provided new and immensely more powerful research tools than existed in earlier centuries, as is obvious by mere reference to electron microscopy in the study of the micro-universe, to the Hubble telescope in the study of the macro-universe, and to the laser, which has become the most powerful research instrument throughout the realm of the science of chemistry. In addition, the laser has found a wide range of uses in medical care.

Finally, since this article was written within easy walking distance of the Stanford Linear Accelerator, it seems appropriate to close with the following observation: in the realm of modern physics it appears that the rate of scientific progress has been largely determined by the availability of improved experimental technologies. In the succinct formulation of Wolfgang Panofsky, the first director of SLAC, "Physics is generally paced by technology and not by the physical laws. We always seem to ask more questions than we have tools to answer." Exactly.

BIBLIOGRAPHY

AARONSON, S. (1979), "The Light of Creation—an Interview with Arno A. Penzias and Robert C. Wilson," *Bell Laboratories Record,* January, p. 13.

BROOKS, H. (1968), "Physics and the Polity", *Science*, vol. 160.

CRICK, Francis (1988), *What Mad Pursuit*, London: Penguin, p. 23.

ELDER, A. L. (ed.) (1970), *The History of Penicillin Production*, New York: American Institute of Chemical Engineers.

GELIJNS, Annetine, and Nathan ROSENBERG (2000), "Diagnostic Devices: An Analysis of Comparative Advantages", in D. Mowery and R. Nelson (eds.), *Sources of Industrial Leadership, Journal of Economic Literature*, 38, chapter 8.

HERRING, unpublished manuscript, n.d.

JUDSON, H. (1979), *The Eighth Day of Creation*, New York: Simon and Schuster.

KRUYTBOSCH, C. (1997), "The Role of Instrumentation in Advancing the Frontiers of Science", in John Irvine *et al., Equipping Science for the 21st Century*, Edward Elgar, chapter 2.

MOWERY, D., and N. ROSENBERG (1998), *Paths of Innovation: Technological Change in 20th Century America,* Cambridge: Cambridge University Press

ROSENBERG, N. (1997), "The Economic Impact of Scientific Instrumentation Developed in Academic Laboratories", in John Irvine *et al., Equipping Science for the 21st Century*, Edward Elgar.

ROSENBERG, N. (1998) "Technological Change in Chemicals: The Role of University-Industry Relations," in A. Arora, R. Landau and N. Rosenberg (eds.), *Chemicals and Long-term Economic Growth*, Wiley.

ROSENBERG, N. (1990), "Why do firms do Basic Research?" *Research Policy* 19(2), pp. 165-174.

ROSENBERG, N. (2002) "America's University/Industry Interfaces, 1945-2000", unpublished manuscript, May.

WATSON, J. (1998), *The Double Helix*, New York: Simon Schuster.

Two Knowledge Dynamics for Innovation

Hiroyuki Itami
Graduate School of Management of Science and Technology
Tokyo University of Science

1. INTRODUCTION

Innovation is the engine of progress in our society. We can define innovation as "dramatically changing people's lives through the introduction of new products or services". Thus, innovation is not a mere technological discovery nor an experimentation of a new product idea. Unless it impacts people's lives, its social significance is minimal. Only when dramatic changes occur in people's lives, does it deserve to be called innovation.

For innovation to be realized in the form of new products or services, two knowledge dynamics are necessary—i.e., knowledge-accumulation dynamics and knowledge-utilization dynamics. Corporations usually introduce new products or services to society and in so doing they utilize various types of knowledge, such as technology and other intangibles which have been created and accumulated by themselves and others in society—e.g. in universities. Thus, for us to understand innovation in our society, we need to understand two dynamics concerning knowledge: how knowledge was accumulated and how knowledge is used. That is the topic of this essay.

Our two main conclusions are: first, organizations are good at accumulating knowledge, and markets are good at using knowledge and second, too much emphasis on the market mechanism can be detrimental to the continuation of innovation since there must be someone who accumulates knowledge in the first place.

2. A TALE OF THREE INNOVATORS: APPLE, MICROSOFT AND GOOGLE

Nowadays, the pace of innovation worldwide is accelerating. For example, many amazing innovations have been achieved in connection with the personal computer over the past three decades, truly changing our lives dramatically. Among the firms which have led this innovation, all of us know that three innovators stand out: Apple, Microsoft and Google. Apple introduced one of the first commercially successful personal computers into our lives and then led the innovation of the user-friendly PC with a mouse and icons. It is now changing the way we read books, through iPad. Microsoft is another innovator in personal computer software, introducing the first widely-used operating system for the Intel chip PC, DOS, and then the user-friendly

Windows operating system, with huge success. Indeed, it is not too much to say that these two firms have been writing the history of personal computers.

After the PC became everybody's tool, the arrival of the internet age led to communication between them, and it brought the famous Google, which made the PC the gateway to the wide world of information on the internet. Google's innovation was the fabulous search engine and super-powerful data centers with innumerable servers which search the web unceasingly. These data centers are now the source of the next generation of computer-related innovation: cloud computing.

All three firms started as small ventures in the US not too long ago, each led by a quintessential entrepreneur—Steve Jobs at Apple, Bill Gates at Microsoft, and Sergey Brin (together with Larry Page) at Google. However, these individuals did not fight single-handedly to make their innovations possible. Behind their entrepreneurial activity lies a huge amount of knowledge accumulated by large organizations. For Jobs and Gates, the large organization was Xerox Corporation, and for Brin it was Stanford University.

The basic technology for the user-friendly personal computer that we know today was developed by Xerox Corporation's Palo Alto Research Center (PARC). This technology was first applied in a workstation called "Alto", which was the precursor of the modern personal computer. However, despite the technical success of the technology, Xerox failed to allocate sufficient resources to this project as a result of various administrative troubles within the company and the poor financial success predicted to the management by the corporate-marketing and accounting departments. Thus, Xerox unfortunately failed to become the market innovator for the personal computer. Even though the technology developed for the personal computer was successful as a result of the large resources devoted to it by Xerox, the accumulated knowledge was not used well by them for their own commercial success.

Feeling disappointed, many engineers left PARC. One group went to Steve Jobs of Apple and created Lisa and Macintosh, the forerunner of today's personal computers. Another group of engineers from PARC was invited by Bill Gates to join Microsoft, where they developed the Windows operating system. Thus, Xerox accumulated most of the necessary basic knowledge for the personal computer age at PARC, but this was later used for commercialization by the small venture firms in Silicon Valley. Entrepreneurs with an acute business sense detected the potential of the technology accumulated by large organizations and capitalized on it by appropriating it through the market mechanism.

For Sergey Brin, the co-founder of Google, the main large organization on whose knowledge accumulation he was able to capitalize was Stanford University. Wikipedia describes his life story until he founded Google as follows:

> Brin immigrated to the United States from the Soviet Union at the age of six. Earning his undergraduate degree at the University of Maryland, he followed in his father's and grandfather's footsteps by studying mathematics, double-majoring in computer science. After graduation, he moved to Stanford to acquire a Ph.D in computer science. There he met Larry Page, whom he quickly befriended. They crammed their dormitory room with inexpensive computers and applied Brin's data mining system to

build a superior search engine. The program became popular at Stanford and they suspended their Ph.D studies to start up Google in a rented garage.

Education at Stanford and the network of professors and students there provided both the accumulated knowledge he could use and the seedbed of ideas and feedbacks for the technological development of the Google search engine.

3. TWO KNOWLEDGE DYNAMICS BY TWO DIFFERENT GROUPS?

The tale of three innovators shows that for an innovation to succeed, the two functions of knowledge about innovation (accumulation of the necessary knowledge and the use of knowledge), are often performed by two separate groups of people or organizations. There are those who accumulate knowledge and others who use it. Of course, there may be happy cases where the same individuals or organizations both accumulate and use, but that is an exception rather than the rule.

Why so? The reasons seem to lie in the nature of the innovation process itself. Usually innovation is a long process and has to go through three very different stages. First, there is a technology-nurturing stage. A new technology is nurtured and developed, using a variety of knowledge that has been accumulated. Second, the new technology must find an entry point into the market in the form of a new product. Let us call this second stage the market-entry stage. Third, the new product introduced must be accepted by a large number of people so that a small entry into the market will become a huge outpouring of demand. Only when this outpouring occurs, will the new product actually be used by many people and thus change their lives. In a sense, society agrees and moves with the new product. Let us call the third stage the societal moving stage. Only when the third stage is successful does innovation finally become a reality.

Knowledge plays an essential role at every stage of innovation. In the first stage, the technological nurturing stage, technological knowledge has to be created and accumulated to make the new technology applicable in reality. Knowledge accumulation dynamics is the key to the first stage. Note that we include knowledge creation in knowledge-accumulation dynamics, since knowledge accumulation presupposes creation of new knowledge to be accumulated. During the second stage of innovation, the market-entry stage, knowledge utilization dynamics is the main knowledge activity. Here, not only the technological knowledge that has been accumulated during the first stage, but also market knowledge is necessary in order to develop a new product. During the third stage of innovation, the societal moving stage, large-scale knowledge diffusion has to occur so that many people may recognize the new product and be interested in trying it out. This is a kind of knowledge use in that many people end up sharing the knowledge about the new product, and the result of knowledge use is its wide dissemination. Thus, knowledge use is the central activity in the last two stages of the innovation process.

Although we categorize the knowledge-relating activity into knowledge-accumulation process and knowledge-utilization process here, each process is a very dynamic and closely-linked activity. For example, knowledge-accumulation itself includes some form of utilization of old knowledge to create new knowledge, and then the combined total of all knowledge, old and new, is accumulated. On the other hand, in the process of knowledge-utilization, a situation is often faced in which the knowledge one starts to use is not enough to develop a new product for market entry, and therefore it is necessary to create new knowledge in order to fill the gap. The newly created knowledge will not disappear after it is used. It will certainly be accumulated in some way after creation. In this sense, knowledge use may be the beginning of another round of knowledge accumulation.

Viewed this way, we can find at least two reasons why two separate groups are often necessary for the entire process of innovation to be successful. One reason is that it is usually a long process from the beginning of an innovation—the technology nurturing stage, till the end—the societal moving stage. Since it is such a long process, a single group of people often cannot last the entire process alone and different groups have to take turns and pass on the baton. Another reason is the difference between knowledge accumulation and knowledge utilization. Those who are good at accumulating knowledge may not be good at using it in the market place.

Even though knowledge accumulation and knowledge utilization are intertwined, it is still meaningful to conceptualize both knowledge-accumulation dynamics and knowledge-utilization dynamics as two separate concepts. Our main point here is that, in order for innovation to occur, both knowledge-accumulation dynamics and knowledge-utilization dynamics are necessary and responsibility for these two dynamics often lies with two different groups of people or in two different places.

4. ORGANIZATIONS ACCUMULATE AND MARKETS UTILIZE

The tale of three innovators also tells us that knowledge-accumulation dynamics often occurs in large organizations, like Xerox and Stanford University, while knowledge is used by the entrepreneurs like Jobs, Gates and Brin, in the market place. It transpires that organization accumulates knowledge and then market uses the accumulated knowledge.

Organizations are the places where people get together and form teams, building a stable human network. In that network, people learn and accumulate together and from each other. Organizations are good at knowledge accumulation. However, they are often not very good at using the accumulation themselves, as in the case of Xerox. Organizations, both corporate and non-corporate, have hierarchical mechanisms of decision-making and resource allocation within their boundaries. This hierarchy often becomes an obstacle to free entrepreneurial experimentation inside the organizations, especially when it involves a large amount of investment. But that kind of investment is inevitable in the second stage of innovation, the market-entry stage. The amount of money involved often becomes huge, whereas at the technology-nurturing stage it is much smaller. The organizational hierarchy is generally not good at selecting the worthwhile risks and thus is not so good at knowledge utilization.

Xerox's failure to invest in the personal computer is an example of such hierarchical failure. Xerox is not, however, an abnormal example. Many large IT firms, like IBM and ATT (American Telephone and Telegraph) accumulated much of the basic knowledge that we have today of IT and communication technology in their labs, at Watson Research Center for IBM and Bell Laboratory for ATT. They were unable, however, to realize the full business potential of their knowledge. Entrepreneurs and the spin-offs from these organizations, like Steve Jobs at Apple, Bill

Gates at Microsoft, Scott McNealy at Sun Micro Systems, Larry Ellison at Oracle, and John Chambers at Cisco Systems, reaped the huge economic benefits from the knowledge that IBM and ATT had accumulated.

Moreover, those entrepreneurs who succeeded in the final two stages of innovation, the market entry stage and the societal moving stage, were often former employees of these big firms. Among the entrepreneurs I have mentioned here, McNealy, Ellison and Chambers all worked once either for IBM or ATT. Only Jobs and Gates were entrepreneurs from the beginning. Ironically, IBM and ATT not only contributed greatly to basic knowledge accumulation for today's IT industry, but also supplied many entrepreneurs who completed innovation in this industry.

Restructuring at IBM and ATT in the 1980's under anti-trust pressure from the US Government was the major trigger for these entrepreneurs to spin off. Thus, the IT revolution in the US would have been impossible without the basic accumulation in those large firms. But it would also have been impossible without the restructuring of these firms who not only released many future entrepreneurs into the new business market but also supplied many engineers who had to leave these big organizations and enter the labor market.

A significant advantage of the market in knowledge utilization comes from its ability to broaden the possibility of combining different items of knowledge accumulated in different organizations across the organizational boundaries. Entrepreneurs are not constrained by organizational boundaries and they are also free from hierarchical control. When they detect an innovation opportunity, they can obtain accumulated knowledge from a large organization by recruiting talent from it or by learning themselves through working for it. Resources can be recombined and reallocated as a result of intelligent moves, thereby enabling the opportunities envisioned to become a reality. The market works as the place for experiment. However, the market is not very good at knowledge accumulation. For knowledge accumulation, a stable human network is needed, like a team, in which learning takes place among its many members. The market is not easily equipped with the capacity to foster such a stable human network since freedom of action by market participants, and in particular the freedom of entry and exit, is the basic principle of market transactions.

In any market economy, we have corporate organizations as its main economic players and we also have non-corporate organizations like universities which specialize in knowledge accumulation. These organizations are linked together by the markets through market transactions to complement their division of labor. Markets also link corporate organizations to consumers. Thus, organizations and markets are two very basic units in any market economy, be it a national or a regional economy.

If we imagine the total picture of knowledge-accumulation dynamics and knowledge- utilization dynamics in the entire economy, our discussion so far implies that organizations, both corporate and non-corporate, function as the main arena for accumulation dynamics. Organizations are the places where accumulation occurs. For utilization dynamics, however, the main arena is the market. The market is where utilization occurs. Obviously, the person who uses accumulated knowledge for the purpose of innovation is the entrepreneur and the corporate organization which he or she leads. That utilization, however, occurs in the context of the market. In short, "organizations accumulate and markets utilize", or a little more precisely, the market allows a firm to utilize the organizations' accumulation.

Certainly in the tale of three innovators and in many other cases, the market mechanism makes it possible for the entrepreneurs to use knowledge accumulated somewhere in the economy. However, it is important to note that there must be someone who accumulates knowledge in the first place. There is no utilization of knowledge unless it is accumulated beforehand.

To summarize, knowledge accumulation is done through learning by a team of people who share a common goal and a common knowledge base. An organization is good at this. Knowledge utilization for innovation is done by experimenting with a new combination of knowledge across organizational boundaries and providing this combination with the right resources at the right time. Entrepreneurs with outstanding ideas are the ones who use this knowledge. The market is the place for this kind of experiment to happen.

5. THE US EXPERIMENTS AND JAPAN DEVELOPS

Although there are corporate organizations and a market in any market economy, the ways they actually work and their relative share of importance in the total economy vary from country to country. Any market economy is a mixture of the organizational (hierarchical) resource-allocation mechanism and the market mechanism.

Within a corporate organization, resources are allocated to the organizational members through the organizational mechanism of hierarchical authority and coordination. In the market place, the market mechanism of competition and price regulates the demand-supply relationship and allocates resources among the market participants. Both mechanisms can vary from time to time and from one country to another in their actual details and the basic patterns of behavior of the participating actors.

If we compare the US and Japan in terms of the mixture of the organization and the market in an economy, much research and the stylized facts seem to indicate that Japan is a more organization-oriented market economy than the US, while the US is more market oriented. For example, in the Japanese intermediate goods market, buyers and sellers tend to maintain long-term transaction relationships and often cooperate long term in innovation. The relationship in automobile-parts transactions in the Japanese automobile industry, often referred to as the Keiretsu relationship, is a typical example, whereas in the US auto parts market, shorter-term, arm's length relationships are the norm. I once termed the pattern of market mechanism with these close relationships the 'organizational' market (a kind of market mechanism with certain features of the organizational mechanism), compared with the freer competitive markets in the US.

If so, it then follows that Japan is better at knowledge-accumulation dynamics and the US is better at knowledge-utilization dynamics. Certainly this seems to be the case and one example is the different nature of innovation activities in the two countries. The US is the country of industrial experiment and Japan is the country of industrial nurturing or development.

The US is well suited for experimental activity aimed at starting up and trying out new businesses or new business models in many industries., Both the capital market and the labor market in the US have the mobility to supply resources for such experiments, and there is ample venture capital as well as a large public-offering market for new companies. Attracted by these markets, both capital and labor flow into the US from all over the world—for example into Silicon Valley.

Over the course of history, there have been many times when the US played a very dominant role in the early stages of commercialization of innovation. Even if we limit our scope to the last forty years, when Japan came to approach the US in industrial strength, the US led the world in semiconductors, liquid crystal display, information technology and biotechnology, among others.

Japan was not too far behind when it came to the development of an industry after the seed was planted. Both in semiconductors and liquid crystal displays, Japan has led the world at various stages of industrial development after the initial experimental stage was over and the pace of technological innovation was accelerating. Another historical example is automobiles. Japan has been overtaking the US as the main player in this industry after the US established it so many years ago. In this process, the Japanese Keiretsu system of inter-firm cooperation, a kind of 'organizational market', played an indispensable role.

How can the US maintain its position? The American knowledge-utilization dynamics seems to be as strong as ever, even now—Google being one of the latest examples. As I noted before, any knowledge utilization presupposes knowledge accumulation. Without accumulation, there is nothing to be utilized. Then, where does knowledge accumulation come from for the American knowledge-utilization dynamics? Perhaps not so much from the American corporate organizations' accumulation as before. For example, the famous Bell Lab disappeared after ATT was broken up and at the IBM Watson Research Center, the glory of its former days is said to be fading.

There are at least two sources of knowledge accumulation still available to American firms and American entrepreneurs. One is the open knowledge base accumulated in American universities. The other source is knowledge accumulation done in other countries, both in corporate organizations and non-corporate organizations. The US can tap and attract those accumulation sources outside its national border.

6. THE US AS THE MARKET ARENA FOR THE WORLD

A particular strength of the American economic system lies in the very openness of this system. One of the clear ways for American firms to capitalize on this openness is to broaden the scope of the open knowledge base they can tap. American firms have been much more active in international sourcing of their knowledge base, either in the form of foreign R&D activities or of inviting foreign university personnel to different American organizations, universities or firms. This is in a sense an effort to broaden the open-knowledge base for American firms.

Another example of broadening the open-knowledge base is to have marketplaces of venture activities for innovation, such as Silicon Valley, in the US. These marketplaces attract many people from all over the world, who bring their accumulated knowledge with them. People from many corners of the world flock to America in order to capitalize on the knowledge accumulation in the market arena that the US provides. In a sense, the US is tapping the large open-knowledge base throughout the world by providing the market arena for knowledge holders outside America.

As we already mentioned, this is made possible partly because of the existence of a very mobile labor market and a very active venture-capital market in the US. There are, however, three more basic conditions that enable the US to function as the market arena for the world.

The first condition is that the native language of the US, English, is the lingua franca of the world. People from other parts of the world can come to the US without fearing language problems, as long as they can speak at least broken English. English has become the lingua franca thanks to the British Empire. The second condition is that the American dollar is the international key currency. People who earn money using the American market system have not had to worry too much about the international value of what they earn, at least until the Lehman shock. The third condition is that the US is a country of immigrants not only in its origin but also in terms of the current immigration policy. The US is thus a melting pot of many people with different ethnic origins where anybody can come from different parts of the world. Those who come to the US do not have to worry too much about their origin.

In a sense, Google is a good example of the US attracting knowledge originally accumulated elsewhere in the world. Sergey Brin came to the US at the age of six from the then Soviet Union, where his father was a mathematics professor at one of the major universities. Sergey followed the family tradition and was educated by his father to become another mathematician, in a sense utilizing the knowledge accumulated in the Soviet Union university system. The rest of the story is now history.

These three conditions—i.e. language, currency and ethnic origin, are something that no other country can currently emulate. Only the US enjoys this special position resulting from its historic and ethnic circumstances. That is why it is able to maintain its knowledge utilization dynamics.

7. FAIR EMPHASIS ON ORGANIZATIONAL KNOWLEDGE ACCUMULATION

In a sense, the US is unique: an exception rather than a rule. If other countries were to try to repeat the glamour of American-style knowledge-utilization dynamics, by trying, for instance, to develop a Silicon Valley of their own without making a substantial effort to accumulate knowledge within their own national borders, they would likely fail. Knowledge utilization does not work without knowledge accumulation in the first place.

Behind the knowledge utilization dynamics in the US lies the very active market mechanism. Economists tend to overemphasize the merits of the market mechanism. But after the fall of communism and the planned economy in the 1990s, the American ideology seems to have swept the world.

It is one thing to use the market mechanism to allocate resources in a stable economy where the knowledge or technology bases do not change very much. The basic theory of a market economy almost always assumes a given set of technology. It is another thing, however, to have too much faith in the market mechanism when we have to consider how to broaden our knowledge base in society by accumulating new knowledge, as in the case of innovation. Who would want to accumulate knowledge if too many economic actors were busy trying to use what they already know?

Moreover, when utilization dynamics become bigger, accumulation dynamics may get smaller. The utilization dynamics would become more active (i.e., get bigger) if there were sources of knowledge that corporations could depend on for utilization. It is often the open knowledge base outside the corporate organizations upon which they become dependent. Such an increase in external dependence could have a negative impact on the corporate organization's efforts to accumulate internal knowledge, because people in the corporate organization might consider it more profitable to use external knowledge rather than invest in obtaining internal knowledge accumulation. Since corporations play an important part in knowledge accumulation through their internal R&D efforts, an increase in their external dependence would mean that the accumulation dynamics in society as a whole would shrink.

Innovation is essential for economic growth, be it a national or a regional economy. We have to be deeply concerned about the mechanism to make innovation more active in the economy as a whole. As I have been emphasizing, an organization accumulates and the market utilizes. There are currently tendencies to emphasize the importance of the market mechanism to such an extent that the importance of the organizational mechanism is neglected. We have to pay attention to the organizational knowledge-accumulation mechanism in both corporate and non-corporate organizations. Too much market orientation may be detrimental to sustainable innovation, to the economy and to society as a whole.

REFERENCES

Chesbrough, H. (2003), *Open Innovation: The New Imperative for Creating and Profiting from Technology,* Harvard Business School Press.

Cringely, R. (1992), *Accidental Empire: How the Boys of Silicon Valley Make Their Millions, Battle Foreign Competition, and Still Can't Get a Date*, Addison-Wesley.Ð

Imai, K., and H. Itami (1984), "Interpenetration of Market and Organization", *International Journal of Industrial Organization*, December 1984.

Itami, H., and T. Roehl (1987), *Mobilizing Invisible Assets*, Harvard University Press.

Itami, H. (1994), "The "Human-Capitalism" of the Japanese Firm as an Integrated System" in Imai and Komiya (eds.), *Business Enterprise in Japan*, The MIT Press.

Itami, H., K. Kusunoki, T. Numagami, *et al*. (eds.) (2010), *Dynamics of Knowledge, Corporate Systems and Innovation*, Springer.

Williamson, O. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press.

BBVA

Innovation Inside and Outside the Company: How Markets for Technology Encourage Open Innovation

Alfonso Gambardella
Università Bocconi, Milan

1. INTRODUCTION

In recent years considerable attention has been devoted to the phenomenon of "open innovation" (Chesbrough, 2003). In a nutshell, the sources for innovation are no longer largely internal in a firm, but have spread to many *loci* in the outside environment. There are different sources of open innovation. A classical one is knowledge spillovers, which arise when firms can capture knowledge or information "in the air", as Marshall put it. Recently, there has been an upsurge in the so-called "open source" phenomenon whereby knowledge and information are distributed openly by their producers, in a context where the production and distribution of knowledge are governed by well-defined norms (e.g., Lerner and Tirole, 2002). An "old" form of open source is open science, which is again based on clear norms of production and diffusion of knowledge (Dasgupta and David, 1994). Open science, and particularly the proximity of firms to universities or other scientific institutions, have themselves been considered sources of knowledge spillovers (e.g., Alcacer and Chung, 2007).

While both spillovers and open source (or open science) are important sources of open innovation, this chapter focuses on the acquisition or distribution of knowledge that depends on a standard economic mechanism—that is, market-mediated forces. While spillovers or open source imply exchange of knowledge based on forces or norms other than markets, in this case knowledge is exchanged at a price. This price may take several forms—e.g., licensing royalties, profit sharing, co-development or supply of resources for innovation. But what distinguishes this source from the others is that knowledge is traded. As we shall see, this trade is more complicated than that of standard commodities, and has far more limitations. However, not only is it possible, but it has become more and more significant in recent years.

We shall discuss the notion of technology trade in a broad sense. While the classical example is licensing agreements, whereby one firm sells technology to another firm in exchange for money, we include more elaborate forms of technology transaction, in particular alliances or other collaborative arrangements for the development of innovation. We remain deliberately vague on this definition to let the reader interpret technology trade in the way he or she prefers.

The reason why we focus on these market-mediated forces is twofold. First, as noted, they have grown in recent years. For example, from 1980 to 2003 in the G8 countries, technology royalty payments and receipts increased annually by an average of 10.7% and reached an annual volume of approximately US$190 billion in 2003 (OECD, 2006). Arora *et al*. (2001), as well as Arora and Gambardella (2010a and 2010b), provide additional systematic evidence of these trends. Second, markets are in general an important institution for economic growth. I am sure that a good deal of knowledge is diffused today via spillovers, open source, or open science, and in this chapter I do not want to discuss or claim any superiority of the market mechanism with respect to these other forces. However, the formation of markets for knowledge, or markets for technology are crucial for many aspects of the growth of knowledge, its diffusion, or the ability of firms to use knowledge as a resource more effectively. Moreover, they create new strategic options for firms, as firms can decide whether to buy, make or sell technology. Without such markets the only strategic option for product innovators is to produce their own technology, and for the technology makers to invest in the downstream assets to sell the product that embodies the technology.

The goal of this chapter is to discuss the factors that make the rise of technology markets possible, along with the limits to their development. I then examine a number of implications for industry structure and company strategy. Specifically, the chapter is organized as follows. The next section provides a general overview of the nature and limits of technology markets. The following section discusses the implications of the absence of markets for assets, highlighting what we can expect to observe when such markets exist. We then deal with three main limitations to the formation of technology markets: a cognitive limitation, i.e., in trading knowledge it may be hard to identify the object of exchange; a transaction cost limitation, i.e., one needs proper institutions for these markets to function; a degree of market limitation—i.e., knowledge has special properties which imply that only under certain conditions does a given piece of knowledge have a market large enough to justify its trade. We conclude by discussing some implications for industry structure and company strategy.

This chapter draws on my research on this topic with Ashish Arora and Andrea Fosfuri. Here I summarize some of the main issues and implications of our work. While they are of course not responsible for any drawbacks or limitations in this chapter, I encourage the interested reader to look at Arora *et al*. (2001a), or Arora and Gambardella (2010a and 2010b), where we discuss at greater length some of the issues dealt with in these pages.

2. BACKGROUND

The exchange of technology between independent parties is not a novel phenomenon. In a series of articles, Lamoreaux and Sokoloff (e.g., 1996 and 1999) discuss the existence of an active market for patents in the US in the 19th century. Typically, inventors used to develop their own technologies which were then sold to firms that developed them and manufactured and commercialized the products, or employed them as process innovation. Interestingly, Lamoreaux and Sokoloff also document the existence of services and institutions

"Even in the mid-20th century many large corporations relied on outside sources of ideas, especially at the upstream end of the knowledge spectrum. For example, Mueller (1962) documents that quite a few of Du Pont's major inventions in the first half of the 20th century came from ideas that it had acquired from external inventors or smaller firms"

in support of the technology trade, as we typically observe when markets exist. Thus, for example, patent lawyers supported both inventors and companies in their trade, special magazines or trade press provided information about the technologies to be sold, and the patent office itself was a crucial institution in this trade as it provided the certification of the property rights and of the novelty of the invention.

As Lamoreaux and Sokoloff themselves noted, the US market for patents shrank in the 1920s. One important reason is that the development of technologies became too risky and complex to be governed by individual inventors. For example, it required expensive equipment that the individual inventors, or their small firms, could not afford. A related reason was that knowledge itself became more complex and interdisciplinary, requiring the contributions of specialists in many fields. As a result, inventors were gradually employed by larger firms, which at the same time started to become large enough to bear the growing costs and risks of innovation.

To be sure, even in the mid-20th century many large corporations relied on outside sources of ideas, especially at the upstream end of the knowledge spectrum. For example, Mueller (1962) documents that quite a few of Du Pont's major inventions in the first half of the 20th century came from ideas that it had acquired from external inventors or smaller firms. Arrow (1983) theorizes on this point. He notes that large and small firms have comparative advantages in different types of innovations or at different stages of the innovation activity. In large firms there is a greater organizational distance between the inventor and the manager who is responsible for financing the innovation. This implies that firms finance only projects in which the asymmetric information between the managers and the inventors is not excessive. But limited asymmetric information is typical of projects for which there is substantial knowledge and information, i.e., that are less risky and innovative. When projects are particularly innovative it is most likely that the inventor has more information than the manager, and in general the manager (or the external financier) takes a greater risk in financing them. Smaller firms, or even firms founded by the inventor, have a shorter organizational distance, which makes the problem of asymmetric information less severe. At the same time, larger firms have more internal resources to finance larger-scale projects and inventions. This is typical of the downstream development of initial innovations or ideas, or of more basic research projects that demand significant investment in large scale equipment or resources. As a result, small firms specialize in the development of riskier projects that demand fewer resources, while larger firms specialize in larger-scale projects, whether upstream or downstream. Given the complementarities between these two types of projects Arrow concludes that a "market for firms", whereby larger firms buy the smaller concerns that produce new ideas, can make our economies more efficient by giving rise to a division of labor in innovation based on comparative advantages.

Teece (1988) analyzes the reasons why a market for Research & Development (R&D) services encounters serious limitations. He argues that the interdependencies among tasks in the innovation process, and the natural uncertainty associated with development and commercialization of innovations, create at least three sources of transaction costs. First, it is hard to provide detailed specifications of the task requirements at the outset of the innovation process. These specifications can be defined more precisely while undertaking processes needing contracts that are largely incomplete and which potentially leave either party open to opportunistic behavior by the other. Second, if a company develops close interactions with one technology supplier, the interplay of relationships may generate sunk costs, which can give rise to switching costs and "lock-in" problems. Finally, releasing pre-contract information to bidders may require the companies to share valuable proprietary information, and increases the risk that competitors will discover its R&D plans.

> "Small firms specialize in the development of riskier projects that demand fewer resources, while larger firms specialize in larger-scale projects, whether upstream or downstream"

Teece (1988) concludes that these reasons explain why the creation, development, and commercialization of new products and processes have traditionally been integrated within firms. This is consistent with the classical view of Grossman and Hart (1986) and Hart and Moore (1990) who argue that vertical integration, and the authority it confers, help to solve the problems of opportunistic behavior that arise when contracts are incomplete. In the case of innovation, this means that a firm can specify and organize the actions of the various agents involved in the innovation process while the process is taking place. In a similar vein, Arrow (1975) develops a model showing that one of the determinants of vertical integration is asymmetric information about the quality of the supply. In addition, being part of the same organization helps the various specialists to acquire a better understanding of each other's problems and needs, to share common objectives and beliefs, and to adopt a common language (Arrow, 1974). This facilitates collaboration and information exchange, and increases the productivity of the innovation process.

Teece (1988) also points out that the problem is more severe in the case of more complex technologies, such as systemic technologies that require profound interdependencies between many activities, as compared to "stand-alone" innovations. He therefore acknowledges that the advantages of integrating the innovation tasks within the same organization can differ across industries and technologies. However, he also provides numerous examples showing how the lack of proper integration of R&D with manufacturing and commercialization leads to poor innovation performance. In short, the Teece (1988) perspective provides the natural theoretical support for the discussion found for instance in Chandler (1990) who argues that historically the advantages of larger integrated firms has been their ability to make systematic "three-pronged" investments in R&D, production and marketing.

> "In order to better understand the implications of markets for technology, it is useful to begin with a more general discussion of the implications of high transaction costs in the markets for corporate assets"

3. THE EFFECTS OF HIGH TRANSACTION COSTS IN THE MARKETS FOR CORPORATE ASSETS

In order to better understand the implications of markets for technology, it is useful to begin with a more general discussion of the implications of high transaction costs in the markets for corporate assets. Broadly speaking, these assets include technology, production expertise and facilities, a strong brand-name reputation, human assets, supplier networks and established marketing channels.

The resource-based theory of the firm suggests that to be a source of sustained above-average performance, resources must meet three criteria: they must be valuable, rare and imperfectly mobile (Barney, 1991; Peteraf, 1993; Markides and Williamson, 1996). In other words, a competitive advantage must be underpinned by resources for which well-functioning markets do not or cannot exist, or they will have high transaction costs. So, the firm builds a sustainable competitive advantage by having access to assets that its competitors cannot access. Barney (1986) notes that the possession of such assets must be rooted in imperfections in the factor market, i.e., the market where the factors used to create such assets are traded, and these imperfections ultimately arise from differences in the expectations that firms hold about the future value of the assets (Barney, 1991). Dierickx and Cool (1989) argue that not all the assets required to sustain competitive advantage can be bought and sold. Instead, such assets must be accumulated internally through a number of mechanisms over a period of time. Similarly, much of the thinking on technology strategy has approached the problem by implicitly or explicitly assuming that technological assets cannot be directly bought and sold and the services of such assets cannot be "rented". In the context of our analysis in this chapter, it is interesting to ask what happens when some assets that were not tradable become tradable.

The immediate consequence of an absent or highly imperfect market for technology is that the innovator has to mainly find the sources of the technology in-house. That is, in order to extract the value from the technology, it (or rather its services) must be embodied in goods and services that are then sold. Such goods and services must have lower costs or command higher prices in order to deliver returns that are greater than the competitive rate of returns: so that firms can earn "quasi-rents".

> "The immediate consequence of an absent or highly imperfect market for technology is that the innovator has to mainly find the sources of the technology in-house"

Consider a case where a firm has developed a new cost-reducing technology for the production of certain goods. In order to extract value from the technology, the firm must use it to produce the goods. Not only does this require the firm to have access to the complementary assets (such as land and physical equipment, marketing channels and so on), but the returns also depend on the volume of output that the firm can produce and sell. If the complementary assets are themselves not traded in a competitive market, or if firms differ in their access to them, then firms that have superior access to these complementary assets will be able to derive greater value from the technology. Similarly, firms that can exploit the technology on a larger scale will be able to derive greater value from it (Cohen and Klepper, 1996; Klepper, 1996).

Continuing with this logic, larger firms or firms with superior access to complementary assets will have a greater incentive to invest in the technology in the first instance. Taking this one step further, firms investing in technology would be well advised to also invest in the complementary assets that cannot be easily and efficiently acquired on the market. In other words, as Teece (1986) puts it, firms have to invest in creating co-specialized assets to maximize their returns from developing new technology. In short, in the absence of a market for technology, a firm must often acquire other assets in order to extract profits from the technology. Insofar as these other assets are themselves expensive and illiquid, well capitalized, large, integrated firms that possess such assets have greater incentives to invest in developing new technologies (Nelson, 1959). Conversely, smaller firms face major hurdles in developing and commercializing technology.

The situation is quite different when the asset can be sold or rented. Complementary assets need not be owned or even directly accessed by the technology developer. The relative importance of complementary assets within the boundaries of the individual firms diminishes relative to the existence of such complementary assets at the level of industries or markets as a whole. Clearly, transaction costs or factors may increase the cost of acquiring the complementary assets externally relative to owning them in-house, even when such markets exist. However, as these imperfections become less important, then, to use Teece's terminology,

> “Consider a case where a firm has developed a new cost-reducing technology for the production of certain goods. In order to extract value from the technology, the firm must use it to produce the goods”

the existence of complementary assets at the level of markets or industries may offset the lack of such assets at the level of the firm.

Ultimately, a market for the asset provides the innovator—a firm that has developed the new technology—with more options. Instead of embodying a newly-developed technology in goods and services, a firm may choose to sell or license it to others, or may choose to buy it from external providers rather than develop it in-house. This does not mean that companies would only acquire technologies from external sources. Leading companies would probably choose the right balance between external acquisition and in-house development of technologies, even though for companies with lower in-house technological capabilities the existence of external technology sources might be critical to enhance their ability to produce and sell more innovative goods. Similarly, a market for technology assets does not mean that innovating firms will become pure licensing companies, although several small (and not so small) firms have been successful as specialized technology suppliers. Rather, the appropriate strategy in the presence of technology markets depends on the efficiency of markets for other types of assets, including finance.

Moreover, in thinking about how a market for technology conditions strategies, there is one other industry level force that must be considered. Markets, particularly efficient markets, are great levelers. As we shall discuss in section 5 below, a technology market lowers entry barriers and increases competition in the product market, which often implies a rethink of existing strategies. In turn, this implies that when a well-functioning market for an asset exists, such an asset cannot be a source of sustainable competitive advantage and firms have to look somewhere else for gaining an edge over competitors. This is an important consequence of technology markets. When they exist, technology cannot be retained in exclusive ways. By contrast, one area that retains its exclusive nature is likely to be the knowledge of customers and markets, and the assets that link these markets and clients to a specific firm vs. its competitors. This knowledge, and the underlying assets depend on sizable investments and a good deal of experience with such markets and clients. To the extent that the markets for these assets remain less perfect than the technology markets, commercialization capabilities, market information, and other downstream assets, may become better sources of differential advantages vis-à-vis the competitors.

4. LIMITS AND OPPORTUNITIES IN THE GROWTH OF TECHNOLOGY MARKETS

4.1. Uncertainty and cognitive limitations

An important limit in the growth of technology markets is that the key objective of exchange, knowledge or technology, is characterized by significant uncertainties. This is especially true when technology is not codified, is embedded in people or machines, and is largely based on experience rather than general principles. For example, improvements in a production process or in a service may be hard to define and codify with precision. In these cases, the object of the transaction is ill-defined to begin with, and this ambiguity makes it harder to trade in the improved process.

Arora and Gambardella (1994) argue that the increase in the extent to which industrial technologies are based in science (including engineering sciences), and the use of advanced instruments and computers, are reducing the fraction of "inarticulable" technology. Thanks to advances in computer technology, including software, many technical problems (e.g., in design, semiconductors, biotechnology, and many other industries) can be defined in logical terms (e.g., mathematical language) and captured in software. Interestingly, there are useful synergies with patents in facilitating technology transactions. Codified technology is easier to patent. Conversely, an increasing appreciation of intellectual property rights encourages codification of innovations.

The difficulties, however, are not only contractual. Discovering who has relevant technology and the price at which they may make it available (if at all) is also difficult. Understanding what they have and how to use it amplifies the problem. Conversely for a seller, identifying potential buyers can be problematic, and once a prospective partner has been identified, settling on the price can be no less challenging. Moreover, new technologies are often surrounded by commercial uncertainty (Rosenberg, 1996). Simply put, it is difficult to know what applications the technology can have. This

"Codified technology is easier to patent. Conversely, an increasing appreciation of intellectual property rights encourages codification of innovations"

raises the search costs of both buyers and suppliers and leads to considerations of option values rather than actual values, and renders potential transactions subject to a variety of biases to which human beings are prone when faced with uncertainty. The net result is that technology transactions are more imperfect and harder to accomplish.

4.2. Patents as a solution to the contractual limitations in technology trade?

Teece's (1988) limitation discussed earlier is fundamentally a limitation in our ability to write contracts involving an ill-defined object like knowledge or technology. Arrow's (1962) solution to the classical problem of information exchange is to appeal for intellectual property protection. If protected, the seller could disclose the details to potential buyers, mitigating the problem. This close relationship between patenting, the technology market, and specialization in invention is reflected in trends in patenting and measures of the technology market. Lamoreaux and Sokoloff note that patenting per capita in America rose during the 19th century, peaked in the early 20th century, and then declined thereafter, closely mirroring

“Many trades in technology actually come with the provision of complementary services—like the provision of know-how and technical services—along with a blueprint technology, like a licensed patent”

trends in individual inventor activities and in trade in patents. After the mid 1980s, patenting per unit of R&D investment in the US changed course and began to rise, very close in time to the resurgence in technology markets as well (see Arora and Gambardella, 2010a).

From a theoretical point of view, Arora (1995) provides a model that clarifies some important conditions under which technology trading can occur, and the role of patents in this process. Many trades in technology actually come with the provision of complementary services—like the provision of know-how and technical services—along with a blueprint technology, like a licensed patent (Taylor and Silberston, 1973, Contractor, 1981). Arora models the case where, along with the technology, the licensor also has to transfer know-how. Given the difficulty in objectively verifying that the know-how is provided, the licensor has an incentive to skimp, since providing such know-how services is costly. Conversely, insofar as some payments are conditional on the provision of the know-how, the licensee has an incentive to withhold payment, claiming that inadequate know-how was provided.

The model shows that these problems can be solved by staggering the payment to the licensor over time and by relying on the property rights of the technology. The buyer’s value depends on the technology and the know-how. While the know-how that is transferred cannot be withdrawn, by withdrawing the rights to use the technology, the licensor does have a hostage because the know-how without a license to the patent is of diminished value. In some cases, the bundling with other complementary inputs, such as specialized machinery can provide a similar role (e.g., Arora, 1996).

The empirical literature provides mixed evidence on the relationship between patent protection and technology licensing contracts. Using a sample of 118 MIT inventions, Gans, Hsu and Stern (2002) find that the presence of patents increases the likelihood that an inventor will license to an incumbent rather than enter the product market by commercializing the invention (see also, Decheneaux *et al*., 2008). Anand and Khanna (2000) find that in the chemicals sector, where patents are believed to be more effective, there are more technology deals, a larger fraction of these are arm’s length, involving exclusive licenses, and a larger fraction of licensing is for future technologies rather than existing ones. In contrast, Cassiman and Veugelers (2002) do not find that more effective patents encourage Belgian firms to enter into collaborative R&D arrangements. Evidence from cross-national data is similarly mixed (see Arora and Gambardella, 2010a)

Arora and Ceccagnoli (2006) provide a potential resolution of this mixed evidence. They argue that when licensing is attractive, then patent protection facilitates licensing. However, for firms with the ability to commercialize technology themselves, patent protection also increases the payoffs to commercialization. Analyzing data from a comprehensive survey of R&D-performing firms in the U.S., they find that patent protection increases licensing, but only for firms that lack complementary manufacturing capabilities. Hall and Ziedonis (2001) provide similar evidence from the semiconductor industry: all else being equal, small design specialists are more likely to patent, and case- study evidence suggests that they do so to license their technologies.

4.3. General-purpose technologies and the size of technology markets

Much of the discussion thus far has focused on the factors that affect the cost and efficiency of technology transactions. Hence Adam Smith's well known observation that "the division of labor is limited by the extent of the market". Thus, even if one could successfully solve the contractual problems, a fully-fledged division of labor in the production and utilization of knowledge and technologies would depend on the size of the market for their applications.

To understand this issue one has to better define what is meant by size of the market in the case of technology. Suppose that a certain body of knowledge or a certain technology is specific to a given application by a particular firm. The context-specific nature of the knowledge and technology would then imply that it is difficult to "re-use" it for other applications. In these cases, the R&D cost can only be spread over the volume of production of the goods associated with the given application. But this implies that the potential supplier would not have any economic advantage in the R&D activity

"A specialization advantage arises only if a supplier which incurs fixed costs can serve a number of different producers at only a small additional cost. This requires the technology or the knowledge base of the supplier not to be totally idiosyncratic to specific contexts or environments"

relative to the firm that produces and sells the goods, because the market size of the technology would not be much larger than that of the goods to which it is applied. Moreover, the comparative advantages of the supplier would not increase, if the size of the market (for those goods, and hence for that application) were to increase. In other words, if a specialized supplier is restricted to a single buyer, there is no advantage to specialization that can offset the inevitable costs of transaction and others involved.

A specialization advantage arises only if a supplier which incurs fixed costs can serve a number of different producers at only a small additional cost. This requires the technology or the knowledge base of the supplier not to be totally idiosyncratic to specific contexts or environments. In other words, while the technology may have to be adapted to

> “Dedicated technologies are customized and co-specialized for a given application, but they cannot be used for other applications”

various applications or users, at least parts of the technology and knowledge bases can be re-used at zero or very low incremental costs. Under these conditions, specialized suppliers would have an advantage over any individual user because although the user could also re-use the knowledge, he or she would do so much less frequently than would a specialized supplier serving a number of users.

In short, what we suggest here is that technology markets and specialized technology suppliers are more likely to arise in the case of general-purpose technologies (Bresnahan and Trajtenberg, 1995; Rosenberg, 1976), or when the technology relies on “general and abstract knowledge bases” (Arora and Gambardella, 1994). General-purpose technologies, or GPTs, are technologies that encompass several applications. Since the fixed cost of developing a GPT can be spread over many potential applications, the efficiency of specialized GPT suppliers increases as the number of applications to which the GPT is applied increases. Thus, specialization advantages arise with increases in the size of the market, insofar as the increase is due to an increase in the number of potential users of the GPT rather than an increase in the size of the individual user or application.

Bresnahan and Gambardella (1998) develop a model in which they argue that the size of the market has two components: N, which is the number of diverse applications of a given technology, and S, the average size of each application. They show that as N increases, a vertical division of labor becomes more likely: technology-specialist firms produce GPTs supplied to the downstream producers that operate in different segments of the final market. As S increases, downstream firms are more likely to integrate backward to produce dedicated technologies for their business. Dedicated technologies are customized and co-specialized for a given application, but they cannot be used for other applications. GPTs can be employed for many applications, but they are less effective than a dedicated technology in any of them.

The intuition is that with high N, a specialized technology supplier can gain economies of scale at the level of the industry. To do so, however, the supplier has to produce a general technology to be offered to the different downstream segments of the final market. By contrast, with high S the final market of a specific application is large enough to justify a fixed cost investment in a technology dedicated to it. Thus, a large N encourages the production of GPTs, and through them the exploitation of economies of scale at the industry level associated with the breadth of the final market. A large S encourages instead the production of dedicated technologies and the exploitation of economies of scale at the level of the firm (or application) associated with the depth of the final market.

Bresnahan and Gambardella discuss the example of the Japanese machine tool sector that developed compact general-purpose machines for the differentiated needs of small final producers in many manufacturing industries. By contrast, in the US, machine tools were technologies largely dedicated to the need of the large automobile industry. The argument is not country-specific, because as they also point out, the roles are reversed in the case of software. In the 1980s, the large PC market in the US, which catered to many different types of users, led to the production of general-purpose packaged software created by specialist vendors. In Japan there were fewer PCs, and computers were largely mainframes owned by large users for specialized applications. As a result, Japan had many large providers of custom software but few independent software package producers. Moreover, users often developed their own software.

Arora *et al*. (2009) test the predictions in Bresnahan and Gambardella using data from the chemical plant engineering sector. In their model, large chemical firms (those investing in more than one plant) choose whether to design the plant internally or engage an external supplier of design and engineering services—the specialized engineering firms, or SEF. Small firms either use an SEF or do not enter the market. They generalize the model by allowing the number of SEFs operating in a market to depend on the demand for their services, and therefore on the decisions of potential buyers, i.e., the chemical firms. Consistent with the theoretical predictions in Bresnahan and Gambardella (1998), they find that the number of SEFs increases when the market expands through an increase in the number of potential buyers but not when market expansion is due to an increase in the average size of buyers.

Today, there are growing examples of GPTs and a rise in related technology markets. Gambardella and McGahan (2010) discuss several of them, and the business models of the technology specialists. For example, Maine and Garsney (2006) discuss the stories of two nanotechnology companies—Hyperion Catalysis and Cambridge Display Technology (CDT). Hyperion Catalysis has developed special applications of fullerenes, carbon allotropes discovered in 1985 that represent a general technology with many potential applications based on basic nanotechnology materials research. Initially Hyperion struggled to find applications for its new materials, and ultimately explored applications via alliances with manufacturers, automotive, aerospace and power generation companies. This has proved a successful strategy, as Hyperion has commercialized more than 40 products in these four distinct markets. CDT has developed polymers that emit light, another general technology with potential applications in semiconductors, consumer electronics and toys. Again, licensing and alliances with several manufacturers have provided the company with paths towards downstream markets.

5. IMPLICATIONS OF MARKET-MEDIATED OPEN INNOVATION PROCESSES

5.1. Efficiency gains from a division of labor in innovation, and a higher rate of innovation

There are many implications of market-mediated open innovation processes. In this concluding section, we discuss three of them: at the level of economies as a whole (this subsection); at the level of industries (the next subsection); and at the level of the individual firms and their strategies.

A technology market creates advantages related to the possibility of creating specialization and division of labor based on

> “In high-tech industries like software, semiconductors, or biotechnology, it is common for new or smaller firms to be set up to explore one innovation. If the innovation fails, the firms can exit the market at small cost, or sell their competencies”

comparative advantages. The argument is the one suggested by Arrow (1983), which we discussed earlier. Different types of firms or agents can specialize in the activities where they are relatively more efficient—e.g., smaller firms in upstream riskier, innovative, projects, and larger firms in downstream or large-scale research projects. The exchange between them creates gains from trade that enhance the overall efficiency of the innovation process.

The gains from trade in technology have three sources. First and foremost, there are advantages related to cost savings for not reinventing the wheel. This aspect is particularly salient in international technology licensing (a country does not have to reinvent what has been invented in another country), and in the discussion of GPTs (an industry may not have to reinvent what has been invented in another industry and can be reused in it). The second source of gains from trade is comparative advantage. Sometimes the inventor of a technology is not the best equipped to develop or market it. Engaging in marketing may even retard innovation, by diverting attention and changing the nature of the organization. Licensing to another firm with a comparative advantage in manufacturing and marketing will yield gains to both parties. The third source of gains is more obvious. For instance, a firm may develop a technology that it does not wish to use but which is applicable elsewhere, and it can gainfully license (or sell) it.

Finally, a division of labor encourages more firms to invest in innovation. To the extent that revenue from innovation is earned only if firms also invest in costly downstream assets, smaller firms, without such capabilities, will on many occasions give up. This is also related to the risks of the innovation process. Low- cost exploration for innovations may entail a small loss in case of failure. The loss will be far more serious if the firm has to invest in marketing capabilities as well. By contrast, firms can try out many innovations if they know that they can sell their intermediate technological outcomes to established firms that already own these assets. In high-tech industries like software, semiconductors, or biotechnology, it is common for new or smaller firms to be set up to explore one innovation. If the innovation fails, the firms can exit the market at small cost, or sell their competencies.

5.2. Implications for entry, competition and industry structure

By definition, open innovation makes technology available more broadly to a larger set of firms. A technology market is an effective mechanism for this process. This reduces entry barriers, increasing competition in product markets. In a world in which the firms have to produce the technology that they use, any ability

> "Technology markets are then the mechanisms that link the growth of first-world markets to the growth of developing-country markets, where the latter is prompted by the higher investment in plants induced by the larger number of SEFs"

or specialization downstream can be undermined by a firm's weakness in developing the technology. By contrast, with technology markets, these firms can use technologies that are developed more efficiently by others, and thus focus on their comparative advantages in the product markets. This raises competition because these firms will be more efficient competitors if they don't have to develop the technology in-house. In some cases, this encourages the entry of firms that would not be able to enter the product markets if they had to develop the technology internally.

The impact of licensing on entry is evident in the chemical industry, which has a long history of licensing of chemical processes (Arora and Gambardella, 1998). Lieberman (1989) finds that licensing was less common in concentrated chemical products, and that when licensing was restricted, there was less entry. In a related study of twenty-four chemical-product markets, Lieberman (1987) reports that patenting by outsiders was associated with a faster decline of product price, once again suggesting that patenting by outsiders encouraged entry into the product market.

Arora *et al*. (2001b) provide more direct evidence in their study of investments in chemical plants in the developing countries during the 1980s. They find that a higher number of SEFs that provide engineering services in product markets is associated with a larger volume of investments in chemical plants by developing-country firms. However, the number of SEFs does not affect investments in developing countries by large multinationals. These multinationals are well-established firms with strong technological capabilities, and thus it is not surprising that they are not affected by the presence of SEFs. By contrast, the domestic chemical firms in the developing countries are technically less advanced but may have an advantage in producing locally. This illustrates the point that technology suppliers—the SEFs in this case—differentially benefit technically less-advanced firms.

This suggests that the division of labor created by the technology markets can be seen as a pecuniary form of externality or spillovers. Spillovers are widely regarded as important in the process of industrial development and economic growth. Yet, much of the research on spillovers focuses on non-pecuniary spillovers, or even more narrowly, on the involuntary (and uncompensated) transfer of knowledge. The SEF story suggests that a market-mediated division of innovative labor can be an important mechanism of knowledge transfer as well. Thus, for example, the growth of the product markets in the developed countries induces the rise of SEFs, which serve the chemical firms in the developing countries. Technology markets are then the mechanisms that link the growth of first-world markets to the growth of developing-country markets, where the latter is prompted by the higher investment in plants induced by the larger number of SEF**s**.

Importantly, these spillovers can also occur across sectors, as our discussion of GPTs suggests. In his study of the US machine tool sector in the 19th century, Rosenberg (1976) noted that the various downstream industries using machine tools started up at different times. For instance, firearm manufacturing emerged earlier than sewing machines, typewriters or bicycles. The growth of the firearm industry spurred the development of metal-cutting and shaping machines. Bicycle production required metal-cutting operations that were very similar to those of the firearm industry (e.g. boring, drilling, milling, planing, grinding, polishing, etc.—see Rosenberg, 1976), and thus the bicycle industry could rely upon the suppliers of metal-cutting machines that were already serving the larger firearm industry. What the suppliers had learned in producing metal-cutting machines for the firearm producers did not have to be learned again to supply bicycles producers. The commonality in the learning process across the industries, or what Rosenberg called "technological convergence", was critical for the transmission of growth, but required the intermediation of an upstream sector.

> "The commonality in the learning process across the industries, or what Rosenberg called "technological convergence", was critical for the transmission of growth, but required the intermediation of an upstream sector"

More generally, an important implication of MFT is that they shift the value of an industry chain downstream. As Dierickx and Cool (1989) suggest, the formation of a market for an asset means that the asset—technology in our case—is no longer strategic to the firm in the sense that it can be used to outcompete others. Markets do not currently exist for assets such as knowledge of markets and customers and for some types of production and distribution assets, and firms can leverage them to obtain a competitive advantage over their rivals. For example, Arora and Gambardella (2005) argue that while innovation in software has grown considerably in emerging economies like India, software innovation in advanced economies continues to derive an important advantage from proximity to lead users. Software innovation depends crucially on close interaction with lead users. These users—e.g., advanced telecom, computer, manufacturing or service firms—are still more unequally distributed worldwide than the capability of producing software innovations.

5.3. Implications for company strategy

Technology markets increase the strategic options of firms. Without them, firms can only make their technologies and use them internally. Now they can buy or make technology, and on the supply side they can profit from their technologies either by using them or by selling them, or both. In Arora and Gambardella (2010a) we discuss at some length how technology markets affect the strategic behavior of companies as technology buyers. Here we focus on their strategic options as technology suppliers.

Arora and Fosfuri (2003) develop a framework to understand the decision of firms to sell technology, and how product-market and technology-market competition condition this decision. In their model, multiple technology holders compete, both in the technology market and in the product market. Technologies are not perfect substitutes for each other, and neither are the goods produced from the technology. In deciding whether to license or not, the technology holder has to balance the revenue from licensing and the rent-dissipation effect produced because licensing will increase product market competition. As a result, factors that enhance licensing revenue or that reduce rent-dissipation will encourage licensing.

This tradeoff depends upon competition in the product market. If the licensee operates in a "distant" market, rent-dissipation is small compared to when the licensee is "nearby". For example, the licensee may operate in a geographical market in which the licensor finds it costly to operate, e.g., because the licensor does not have the complementary downstream assets. Similarly, the technology could be used for a different type of product that the licensor may not produce. Arora and Fosfuri note that product market competition enhances licensing because rent dissipation falls faster than licensing rents as product market competition increases. Indeed, as is well known, a monopolist will not license. Consistent with this, Lieberman (1989) finds that licensing was less common in concentrated chemical products, and the limited licensing that did take place was by outsiders (non producers and foreign firms).

"Product market competition enhances licensing because rent dissipation falls faster than licensing rents as product market competition increases. Indeed, as is well known, a monopolist will not license"

Arora and Fosfuri also point out that licensing is more likely when products are homogeneous rather than differentiated. If products are differentiated, a licensee is closer in the product space to the licensor than to other producers, so that the rent dissipation felt by the licensor is greater than if the product is homogenous. Put differently, by licensing, a technology holder imposes a greater negative (pecuniary) externality on other producers when the product is homogenous. Consistent with this, Fosfuri (2006) finds that licensing is lower in markets where technology-specific product differentiation is high.

The Arora-Fosfuri framework also implies that smaller firms are more likely to license, because they suffer less from the rent-dissipation from additional competitors. The logic is apparent in the extreme case in which the licensor has no stakes in the downstream markets, and thus has no product market rents to worry about. This is also consistent with the observation that in many high-tech industries (e.g., biotechnology, semiconductors, software) suppliers often do not produce in the product markets for which they supply technology, and with the evidence provided by other studies in the literature, as discussed in Arora and Gambardella (2010a).

Gambardella and Giarratana (2010) generalize the Arora and Fosfuri framework by emphasizing the interplay between the generality of the technology and the fragmentation of the product markets. The generality of the technology makes it attractive to “distant” user firms, which implies that rents from licensing can be earned from firms in product markets different from that of the technology holder. Because the markets are distant in product space, the rent-dissipation is small, which raises the incentives to license.

Gambardella and Giarratana (2010) jointly consider both the licensing decision and the decision on the range of product markets that the technology holder will enter. The key assumption is that technology can be deployed in more product markets than is profitable for the technology holder to serve directly, suggesting that technology can have broader economies of scope than marketing and manufacturing assets, which creates opportunities for licensing. In particular, GPTs can be so broadly applicable that few firms are likely to exploit all the applications.

6. CONCLUSIONS

There are several conduits to open innovation: knowledge spillovers, open science, and more recently open source technologies. This chapter focused on a market-mediated mechanism. It shows that a good deal of the open innovation process can take place through this standard economic channel.

The role of technology markets is best appreciated by looking at what happens when they exhibit high transaction costs. The most significant implication is that any user of technology has to possess the resources and capabilities to produce it, and any producer of technology has to have the resources and capabilities to embody it in final products to be commercialized. This has natural constraints in that it foregoes the advantages of specialization and division of labor according to comparative advantages, with implied market-level inefficiencies. In addition, the rate of innovation is reduced because any inventor or technology-specialist firm can invest in innovation only if they also make the far more costly investments in the assets or the capabilities that enable them to enjoy revenue from the commercialization of the final products.

When technology markets function well, and make innovation open, entry and competition increase because technology is no longer the highly-guarded secret of some master firm or R&D lab. An important implication is that technology becomes less strategic as a source of competitive advantage of firms. Firms have to focus on other strategic assets, which are less open and more unique to each of them. Among others, one of such assets is the ability of the firm to secure customers and final markets through investments in downstream assets. This also suggests why many technology specialist firms make little or no revenue (e.g., the biotech firms—see Pisano, 2006) and a good deal of the revenue in the vertical structure of an industry accrues downstream.

From a strategic point of view, technology markets give firms more options. Apart from buying rather than making technology, they can sell their own technologies. This is a strategic choice that hinges on the comparison of the rents from licensing and the dissipation of rents caused by the nurturing of a potential competitor. As technologies become broader, and they have a larger span of applications than the ability of a firm to exploit all of them, this can become an increasingly feasible strategy.

This chapter neglects several important issues. For example, the demand for technology or uncertainty plays an important role in these markets and in the ensuing openness of the innovation process. Some of them are discussed in Arora *et al*. (2001a) and more recently in Arora and Gambardella (2010a and 2010b). The interested reader is welcome to consult these sources, and the many references therein, for further insights.

REFERENCES

ALCACER, J., and W. CHUNG (2007), "Location Strategies and Knowledge Spillovers", *Management Science* 53, pp. 760-776.

ANAND, B., and T. KHANNA (2000), "The Structure of Licensing Contracts", *Journal of Industrial Economics* 48(1), pp.103-135.

ARORA, A. (1995), "Licensing Tacit Knowledge: Intellectual Property Rights and the Market for Know-how", *Economics of Innovation and New Technology* 4, pp. 41-59.

ARORA, A. (1996), "Contracting for Tacit Knowledge: The Provision of Technical Services in Technology Licensing Contracts", *Journal of Development Economics* 50(2), pp. 233-256.

ARORA, A., and M. CECCAGNOLI (2006), "Patent Protection, Complementary Assets, and Firms: Incentives for Technology Licensing", *Management Science* 52, pp. 293-308.

ARORA, A., and A. FOSFURI (2003), "Licensing the Market for Technology", *Journal of Economic Behavior & Organization* 52(2), pp. 277-295.

ARORA, A., A. FOSFURI and A. GAMBARDELLA (2001a), *Markets for Technology: The Economics of Innovation and Corporate Strategy*, Cambridge MA: The MIT Press.

ARORA, A., A. FOSFURI and A. GAMBARDELLA (2001b), "Specialized Technology Suppliers, International Spillovers and Investments: Evidence from the Chemical Industry", *Journal of Development Economics* 65(1), pp. 31-54.

ARORA, A., and A. GAMBARDELLA (1994), "The Changing Technology of Technical Change: General and Abstract Knowledge and the Division of Innovative Labour", *Research Policy* 23, pp. 523-532.

ARORA, A., and A. GAMBARDELLA (1998), "Evolution of Industry Structure in the Chemical Industry", in A. Arora, R. Landau and N. Rosenberg (eds.), *Chemicals and Long-Term Economic Growth*, New York: John Wiley & Sons.

ARORA, A., and A. GAMBARDELLA (2005) ,"The Globalization of the Software Industry: Perspectives and Opportunities for Developed and Developing Countries", in A. Jaffe, J. Lerner and S. Stern (eds.), *Innovation Policy and the Economy* 5, Cambridge MA: The MIT Press.

ARORA, A., and A. GAMBARDELLA (2010a), "Markets for Technology", in B. Hall and N. Rosenberg (eds), *Handbook of Economics of Innovation*, Amsterdam: Elsevier.

ARORA, A., and A. GAMBARDELLA (2010b), "Ideas for Rent: An Overview of Markets for Technology", *Industrial and Corporate Change* 19, pp. 775-803.

ARORA, A., W. VOGT and J. YOON (2009), "Is the Division of Labor Limited by the Extent of the Market? Evidence from the Chemical Industry", *Industrial and Corporate Change* 18(5), pp. 785-806.

ARROW, K. (1962), "Economic Welfare and the Allocation of Resources for Invention", in *The Rate and Direction of Inventive Activity*, Princeton NJ: Princeton University Press, pp. 609-625.

ARROW, K. (1974), *The Limits of Organization*, New York, NY: W.W. Nortong.

ARROW, K. (1975), "Vertical Integration and Communication", *Bell Journal of Economics* 6, pp. 173-183.

ARROW, K. (1983), "Innovation in Large and Small Firms", in J. Ronen (ed.), *Entrepreneurship*, Lexington MA: Lexington Books, pp.15-28.

BARNEY, J. (1986), "Strategic Factor Markets: Expectations, Luck and Business Strategy", *Management Science* 32(10), pp. 1231-1241.

BARNEY, J. (1991), "Firms' Resources and Sustained Competitive Advantage", *Journal of Management* 17, pp. 99-120.

BRESNAHAN, T., and A. GAMBARDELLA (1998), "The Division of Inventive Labor and the Extent of the Market", in E. Helpman (ed.) *General Purpose Technologies and Economic Growth*, Cambridge, MA: The MIT Press,

BRESNAHAN, T., and M. TRAJTENBERG (1995), "General Purpose Technologies: 'Engines of Growth'?", *Journal of Econometrics* 65(1), pp. 83-108.

CASSIMAN, B., and R. VEUGELERS (2002), "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium", *American Economic Review* 92(4), pp. 1169-1184.

CHANDLER, A. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*, Part I: "Introduction: Scale and Scope", Cambridge MA: The Belknap Press of Harvard University Press, pp. 1-46.

CHESBROUGH, H. (2003), *Open innovation: The New Imperative for Creating and Profiting from Technology*, Boston, MA: Harvard Business School Press.
COHEN, W., and S. KLEPPER (1996), "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D", *Review of Economics and Statistics* 78(2), pp. 232-243.
CONTRACTOR, F. (1981), "International Technology Licensing", Lexington, MA: DC Heath and Company.
DASGUPTA, P., and P. DAVID (1994), "Towards a New Economics of Science", *Research Policy* 23, pp. 487-521.
DECHENAUX, E., M. THURSBY and J. THURSBY (2009), "Shirking, Sharing Risk and Shelving: The Role of University License Contracts", *International Journal of Industrial Organization* 27(1), pp. 80-91.
DIERICKX, I., and K. COOL (1989) "Asset Stock Accumulation and Sustainability of Competitive Advantage", *Management Science* 35(12), pp, 1504-1513.
FOSFURI, A. (2006), "The Licensing Dilemma: Understanding the Determinants of the Rate of Technology Licensing", *Strategic Management Journal* 27, pp. 1141-1158.
GAMBARDELLA, A., and M. GIARRATANA (2010), "General Technological Capabilities, Product Market Fragmentation, and the Market for Technology: Evidence from the Software Security Industry", draft, Milan: Bocconi University.
GAMBARDELLA, A., and A. MCGAHAN (2010), "Business-Model Innovation, General Purpose Technologies, Specialization and Industry Change", *Long Range Planning* 43, pp. 262-271.
GANS, J., D. HSU and S. STERN (2002), "When Does Start- Up Innovation Raise the Gale of Creative Destruction", *Rand Journal of Economics* 33 (4), pp. 571-586.
GROSSMAN, S., and O. HART (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy* 94, pp. 691-719.
HALL, B., and R. ZIEDONIS (2001), "The Patent Paradox Revisited: Determinants of Patenting in the US Semiconductor Industry, 1980-1994", *Rand Journal of Economics* 32(1), pp. 101-128.
HART, O., and J. MOORE (1990) "Property Rights and the Nature of the Firm", *Journal of Political Economy* 98, pp. 1119-1158.
KLEPPER, S. (1996), "Entry, Exit, Growth, and Innovation over the Product Life Cycle", *American Economic Review* 86(3), pp. 562-583.
LAMOREAUX, N., and K. SOKOLOFF (1996), "Long-Term Change in the Organization of Inventive Activity", *Proceedings of the National Academy of Sciences of the United States of America* 93(23), pp. 12686-12692.
LAMOREAUX, N., and K. SOKOLOFF (1999), "Inventors, Firms, and the Market for Technology in the late Nineteenth and Early Twentieth Centuries", in N. Lamoreaux, D. Raff and P. Temin (eds.), *Learning by Doing in Markets, Firms, and Countries*, Chicago, IL: University of Chicago Press.
LERNER, J., and J. TIROLE (2002), "Some Simple Economics of Open Source", *Journal of Industrial Economics* L(2), pp. 197-234.
LIEBERMAN, M. (1987), "Patents, Learning by Doing, and Market Structure in the Chemical Processing Industries", *International Journal of Industrial Organization* 5(3), pp. 257-276.
LIEBERMAN, M. (1989),"The Learning Curve, Technological Barriers to Entry, and Competitive Survival in the Chemical-Processing Industries", *Strategic Management Journal* 10(5), pp. 431-447.
MAINE, E., and E. GARSNEY (2006), "Commercializing Generic Technology: The Case of Advanced Materials Ventures", *Research Policy* 35(3), pp. 375-393.
MARKIDES, C., and P. WILLIAMSON (1996) "Corporate Diversification and Organization Structure: A Resource-Based View", *Academy of Management Journal* 39, pp. 340-367.
MUELLER, W. (1962), "The Origins of the Basic Inventions Underlying Du Pont's Major Product and Process Innovations, 1920 to 1950", in *The Rate and Direction of Inventive Activity*, Princeton NJ: Princeton University Press, pp. 323-346.
NELSON, R. (1959), "The Simple Economics of Basic Scientific Research", *Journal of Political Economy* 67(2), pp. 297-306.
OECD (2006), *OECD Technology Indicators,Technology Balance of Payment—Payments/Receipts*, Paris: OECD.
PETERAF, M. (1993) "The Cornerstones of Competitive Advantage: A Resource-Based View", *Strategic Management Journal* 14(3), pp. 179-191.
PISANO, G. (2006), "Can Science Be a Business: Lessons from Biotech", *Harvard Business Review*, October, pp. 114-125.
ROSENBERG, N. (1976), "Technological Change in the Machine Tool Industry" in N. Rosenberg, (ed.) *Perspectives on Technology*, Cambridge MA: Cambridge University Press, pp. 11-31.
ROSENBERG, N. (1996), "Uncertainty and Technological Change" in R. Landau T. Taylor and G. Wright (eds.), *The Mosaic of Economic Growth*, Stanford, CA: Stanford University Press.
TAYLOR, C., and A. SILBERSTON (1973),"The Economic Impact of the Patent System: A Study of the British Experience", Cambridge, MA: University of Cambridge, D.A.E. monograph 23, Cambridge University Press.
TEECE, D. (1986), "Profiting from Technological Innovation", *Research Policy* 15, pp. 285-305.
TEECE, D. (1988), "Technological Change and the Nature of the Firm", in G. Dosi *et al*. (eds.) *Technological Change and Economic Theory*, London: Printer Publishers, pp. 257-281.

BBVA

Culture, values and the long waves of capitalist development

Francisco Louçã
ISEG (Technical University of Lisbon), UECE

'Time present and time past
Are both perhaps present in time future
And time future contained in time past'
T. S. Eliot, *Four Quartets*

1. INTRODUCTION: HOW THE HISTORIES OF ECONOMIES, TECHNOLOGIES AND CULTURES INTERMINGLE

The standard literature has recognized, not without theoretical misapprehension, the imposing recurrence of the short 'Juglar' or business cycles, but instead has generally avoided the longer grasp of history in economic analysis. This chapter argues that understanding the larger processes of social, economic, technological and even cultural innovation in modern economies requires establishing the framework of cultural values, of social relations in production and trade, and of the establishment of institutions and learning processes, and that these require the interpretation of recurrent long waves of capitalist development. Furthermore, it argues that an understanding of the formation of cultures and values can benefit from analysis of the historic framework of the successive modes of development of modern economies -- what have been called the long or Kondratiev waves.

There are two sound reasons for choosing this approach and framework. The first is the crashing evidence of facts: between 2007 and 2009 the developed economies suffered the deepest general recession since 1929, proving the impact of the Juglar, but this happened after decades of mild expansion with many recessions, with low rates of accumulation and deep financial and structural mismatches, evoking the impact of the longer processes of economic and social rearrangements, described by the long waves. The second reason is also relevant for the purpose of this collective book, and it is the evidence of the impact of historical processes such as the technological revolution in course. As Chris Freeman has consistently argued, the crux of the matter in economic development is either the match or the mismatch between the techno-economic and the socio-institutional systems, and these long phases of adjustment or crisis mark each age of modern economic growth, or capitalist development.[1] The question for many is therefore, why is it that the ongoing deep technological revolution is so slow to

[1] This section and the last one are largely based upon the book *As Time Goes By—From the Industrial Revolutions to the Information Revolution*, co-authored with the late Chris Freeman some years ago (Oxford University Press, 2002). Freeman, who died in the summer 2010, was certainly one of the leading researchers on innovation and evolutionary economics.

change the general economic conditions? Or, for the purpose of this chapter, how is it that the changes in the general economic conditions contain or shape the evolution of values and cultures?

In the following paper I will argue that this change is on its way, and that it is deeper than commonly suspected. In particular, the interest is concentrated here on the mutations in the landscape of culture, both in the strictest sense of the production of cultural artefacts generating sense and reference, and in the widest sense on the changes in values spreading over communities and societies under the impact of challenging radical innovation.

In order to present this short contribution, I will summarise the common characteristics of each of the four long waves, also present in the emergence of the probable fifth wave. According to this view, the reason for the long contradictory process of structural adjustment and successive crises with low profitability and accumulation—and strong tendencies to concentrate capital on short-term adventurous financial applications—is that there is a mismatch between the already available technological capabilities and the economic restructuring of the major economies. This was the reason for previous long periods of slow expansion and general crises, and it may be the case nowadays, as described by previous long waves of economic development. Furthermore, this mismatch generates social and cultural tensions.

Whereas, however, many of the earlier long-wave theories relied mainly or exclusively on statistical evidence of fluctuations in rates of growth of GDP, industrial production or prices, Freeman and I argued in *As Time Goes By* that such aggregates conceal as much as they reveal and that the really important long-wave phenomena were the successive structural transformations of the economic system brought about by successive waves of technical change and the accompanying organizational and managerial changes. Gerhard Mensch used the expression 'metamorphoses' to characterize these transformations and this is a good way to describe what has taken place -- a process Schumpeter had emphasised.

From this standpoint it was unfortunate that many of those investigating and developing the long-wave concept followed Kondratiev in attempting to substantiate their ideas with purely statistical evidence of *aggregate* movements in production and prices, rather than evidence of structural transformation and waves of technical and economic change.[2] This made it possible for those who believed that the test of a theory was exclusively in terms of econometric procedures on data and on aggregate trends to mount a plausible attack on the very idea of long waves.

Instead, in our book we challenged our colleagues to consider that, during a period of turbulent structural change, some new industries and activities grow very rapidly but others decline, stagnate or grow more slowly. The combined outcome of these contradictory tendencies will vary in different countries at different times, depending on wider political and institutional factors, as well as on more narrowly defined economic and technological trends. Typically, a structural crisis of adjustment will tend to be a period when the expansionary impetus from emerging constellations of new products, processes and organizational innovations will not yet be widespread enough to overcome the depressive constraining effects of the slow-down or contraction in the older established industries.

[2] In spite of this, I follow the standard procedure introduced by Schumpeter to call these long phases and movements 'Kondratiev waves', since this author was the introducer of the modern debate on the historical trends in capitalist development.

> “Typically, a structural crisis of adjustment will tend to be a period when the expansionary impetus from emerging constellations of new products, processes and organizational innovations will not yet be widespread enough to overcome the depressive constraining effects of the slow-down or contraction in the older established industries”

However, this may not always be the case. The expansionary impetus from the new developments may be so great that it imparts an upward thrust to aggregate industrial production and/or GDP despite a structural crisis of adaptation and high levels of structural unemployment. This was apparently the case in Britain in the 1830s and 1840s and in the United States in the 1880s and 1920s. On the other hand, the tempestuous growth of the automobile and oil industries in the 1920s was not sufficient to overcome the depressive trends in the US and the world economy in the 1930s, exacerbated as they were by severe political crises, international conflicts and monetary crises. This is apparently the case today, with the lasting Solow paradox of general computerization but scarce effects on productivity.

Qualitative historical narrative, as well as disaggregated sectoral data, are more important than aggregate *quantitative* data in analysing successive industrial revolutions. As Keynes pointed out in his debate with Tinbergen, one of the main dangers in the standard statistical procedures is that they may obscure or altogether ignore processes of qualitative change.

However, to justify the use of the concept of ‘waves’ or ‘cycles’, rather than simply ‘stages’ or ‘periods’ of historical evolution, it is necessary to distinguish *recurrent* phenomena in each period as well as the unique features of each technological revolution. Moreover, it is essential to place these recurrent features of the changes in technology and the economy in a wider institutional and social context, a context in which political and cultural changes may sometimes predominate in determining the course of events.

In any case, a theoretical framework for the history of economic growth should satisfy three main requirements. First, it should provide a plausible explanation and illumination of the stylized facts that summarize the main features of the growth of the world economy. This is essential to pave the way for generalizations, which should of course be constantly tested against new historical evidence, as well as newly unfolding events. Secondly, it should do this for the three main categories identified by Abramovitz: forging ahead, catching up and falling behind, in order to discuss the uneven development of different economies. Finally, it should provide a framework for analyzing and reconciling the research data, case studies and generalizations emerging from the various sub-disciplines of history: the history of science and of technology, economic history, political history and cultural history.

As a step in an inevitably ambitious and hazardous undertaking, the following definitions were used in previous work:

1. *The history of science* is the history of those institutions and sub-systems of society which are primarily concerned with the advancement of knowledge about the natural world and the ideas of those individuals (whether working in specialized institutions or not) whose activity is directed towards this objective.

2. *The history of technology* is the history of artefacts and techniques and of the activities of those individuals, groups, institutions and sub-systems of society which are primarily concerned with their design, development and improvement, and with the recording and dissemination of the knowledge used for these activities.

3. *Economic history* is the history of those institutions and sub-systems of society which are primarily concerned with the production, distribution and consumption of goods and services and of those individuals and institutions concerned with the organization of these activities.

4. *Political history* is the history of those individuals, institutions and sub-systems of society which are primarily concerned with the governance (legal and political regulation by central, local or international authorities) of society, including its military affairs.

5. *Cultural history* is the history of those ideas, values, artistic creations, traditions, religions and customs which influence the behavioural norms of society and of those individuals and institutions which promote them. The next section will present an overview of the major changes in cultural history for the last decades, relating such processes to the major social, economic and technological innovations.

This chapter will refer to these five sub-divisions for conceptual and analytical purposes, whilst accepting of course that people make only one history and recognizing that in real life the five streams intermingle. However, the use of sub-divisions is *not* simply a matter of convenience in handling an extremely complicated topic, nor is it just a question of following the academic departmentalization and specializations which have emerged in the twentieth century and that were even accentuated in this century. Moreover, the establishment of separate sub-disciplines reflects the sense of dissatisfaction felt especially by scientists, technologists and economists that their special interests were being neglected within the wider rubric in which they were contained. Some protested against the neglect of technology in this approach, and I will add that other factors are also relevant, such as a wider range of cultural phenomena, in order to understand the reluctance to accept new technologies, the social imbalances these create, and the deep changes created by such innovation, including in the formation of novel values.

These five sub-divisions are proposed for fundamental reasons. In the first place, they are proposed because each one has been shown to have some independent influence on the process of economic growth, varying in different periods and different parts of the world, but at least sometimes extending over long periods. Finally, and most important of all, it is precisely the *relative* autonomy of each of these five processes which can give rise to problems of lack of synchronicity and harmony or, alternatively, of harmonious integration and virtuous circle effects on economic growth. It is thus essential to study both the relatively independent development of each stream of history and their interdependencies, their loss of integration and their reintegration, for the understanding of the long waves in the history of capitalism depends on these movements of synchronicity and mismatch.[3]

The next section summarises how these movements proceed, presenting some conclusions on the recurrence of economic and technological processes that account for the long waves, while the following one addresses the impact of systemic changes on culture. Finally, some conclusions on the contemporary problems of economic development are presented.

2. THE LONG WAVES AS THE RESULT OF FIVE RECURRING PROCESSES IN ECONOMIC HISTORY

An historical approach to economic growth is unlikely to be acceptable, unless it not only tells a story using this type of theoretical framework, but is capable of identifying and explaining *recurrent* phenomena, as well as special cases. As Werner Sombart (1929) put it, 'all history and particularly economic history has to deal not only or mainly with the special case, but with events and situations which recur, and, recurring, exhibit some similarity of feature—instances which can be grouped together, given a collective label and treated as a whole' (Sombart, 1929: 18).

For this purpose, five recurrent processes involving the shaping of economic development were identified in our previous work: the creation of super profits of innovative entrepreneurship, the pervasive constellations of technological innovations, the organizational and management changes as the result of such impulses, the general crises of structural adjustment and those of the regulatory regimes. Together, they account for the existence of long waves as recurrent phenomena in modern economic development.

2.1. The recurrence of exceptional super-profits of innovative entrepreneurship in successive long waves

Both some of the sternest critics of capitalism (for example, Karl Marx) and some of its most ardent admirers (for example, Friedrich von Hayek) have argued that one of the foremost characteristics of capitalism has been its capacity to generate and to diffuse a torrent of technical innovations. Everything that is solid melts into air, claimed Marx in order to descry the hurricane of innovation in modern times.

The exceptionally favourable confluence of cultural, political, economic, geographical, scientific and social circumstances in eighteenth century Britain gave rise to that upsurge of technical and organizational innovations known ever since as the 'Industrial Revolution'. It is also understood that other capitalist economies, and especially that of the United States, were not only able to achieve similar results but, as time went by, were also able to outstrip Britain with new constellations of innovations, namely within the framework of the second technological revolution and of the age of electricity and the automobile.

Capitalist economies have been able to achieve these remarkable results, 'surpassing the wonders of the Ancient World', as Marx and Engels again put it, by a combination of incentives and pressures ultimately affecting numerous firms and individuals: in short, they were able to do so through a culture of innovation and organization of structural change. First of all, of course, a well-functioning capitalist economy offers

[3] This deals with developments within *industrial capitalist* economies, and does not address other issues. Indeed, there are other types of theories of long cycles which have a far wider scope, even going back to the Ancient World. Instead, the currently summarized theoretical sketch has a relatively limited domain of application: it relates to the evolution of capitalist economies from the late eighteenth to the early twenty-first centuries and it postulates for this period the predominance within the leading economies of recognizably capitalist institutions and in particular of private ownership and private wealth accumulation through profits.
To criticize this theory as 'technological determinism' is therefore wide of the mark. It is the very existence of certain social institutions which made possible the technological revolutions which have been shortly described. Moreover, these successive new technologies discussed here were not 'manna from heaven'; they were the outcome of human social activities and institutions. Within this general framework, giving emphasis to the changes in technology as a dynamic element in the whole system is simply a way to stress crucial changes moving the whole economic and social process.

the possibility, but by no means the certainty, of *profit* from successful innovation, and sometimes very large profit. This profit may be accompanied by other rewards: status, privilege, political advancement and fame. Some of the most successful entrepreneurs in each technological revolution did indeed achieve extraordinarily large profits, although they did not necessarily seek the other advantages often sought by very wealthy individuals. Fame itself they could hardly avoid and indeed this was a very important social mechanism for the dissemination of their innovations and for efforts to surpass them. Arkwright, Wedgwood, Hudson, Brunel, the Vanderbilts, Carnegie, Krupp, Rockefeller, Rathenau, Siemens, Diesel, Ford, Gates and Murdoch are all examples which we have cited of entrepreneurs and inventors, who achieved both fame and fortune through their innovations, whether technical, organizational or both. Schumpeter emphasised this trend of entrepreneurship, moved by 'social deviants' breaking the routine. In short, capitalism is adaptive since it rejects equilibrium.

A number of long-wave theorists (Mandel, 1980; Goodwin, 1985; Poletayev, 1987) have constructed models of the behaviour of the economic system based mainly on long-term fluctuations in the *aggregate* rate of profit. They have argued quite plausibly that a fall in the rate of profit tends to occur after a long period of prosperity and expansion, partly because of the Schumpeterian processes of erosion of innovators' profits during dissemination and partly on account of wider pressures from rising costs of inputs. These tendencies for the rate of profit to fall at the peak of a long boom are among the main reasons explaining the upper turning point in the long wave and the onset of a prolonged downswing in which generally lower rates of profit prevail.

The statistics are very difficult to assemble, especially for the nineteenth century but, such as they are, they do provide some support for this interpretation (e.g. Entor and Poletayev, 1985). The plausibility of these models cannot be denied, but since the concern of this chapter is mainly on *structural* change, it is more accurate to stress here the exceptionally large 'super-profits' which may be realized through the exploitation of major radical innovations. These profits appear all the more remarkable if they are made during a period of general decline in the rate of profit in the 'downswing' phase of the long wave. Although he disagrees with Mandel and other long-wave theorists on the *aggregate* rate of profit, Tylecote (1992) also points to the extraordinary importance of the demonstration effect for key innovations in each long wave.

This demonstration effect is not only one of clear-cut technical efficiency but also one of great profitability and great potential for widespread application. This effect was so powerful in the case of Arkwright's water-frame that it led some of his rivals and competitors to try to destroy his equipment. Despite this hostility, the successful and highly profitable operations of Cromford mill and his other factories stimulated numerous imitators to invest in cotton mills, especially after the expiry of his disputed patents. Some of the early canal investments, such as the Worsley-Manchester Canal, made very good profits. On a far greater scale, the Rainhill Trials of various steam locomotives, followed by the successful and profitable operation of the Liverpool-Manchester Railway, led to an enormous boom in railway investment and indeed to a huge financial bubble based on the excitement caused by often exaggerated estimates of the potential profits to be made.

Railway promoters, such as George Hudson in Britain and the Vanderbilts in the United States, made huge profits from speculation and financial manipulation, rather than technical innovation, even though Hudson lost his fortune in the end. Otherwise, the profits of Carnegie, Krupp and Ford provided examples of the vast amounts that could be accumulated by successful innovative entrepreneurship. The profits of IBM were not so much the result of *individual* entrepreneurship as of *company* performance; they were nevertheless hugely impressive and IBM was in some measure the most profitable firm in the world before it suffered setbacks in the 1980s, and its place as the most profitable player in ICT was usurped by Microsoft and, currently, by a number of challenging firms investing in the interface between mobile communications and internet.

The first distinguishing recurrent characteristic of the long waves, therefore, is that in each case, although the individual innovations were unique and very different, a cluster of innovations emerged which offered a clear-cut potential for immense profits, based on proven technical superiority over previous modes of production. Minor incremental improvements were, of course, occurring all the time, but the innovations which were at the heart of each wave offered quite dramatic changes in productivity and profitability. However, these highly profitable innovations were not isolated events but part of a constellation of inter-related product, process and organizational innovations. Numerous other firms jumped on the band-wagon, as Schumpeter had suggested, including many small new firms. Sometimes it was a new *process*, which generated the main super-profits, sometimes it was an array of new products, sometimes it was mainly organizational changes, as in the case of Ford's assembly line or the internet, but in all cases there were interdependent developments, both technically and economically. The dramatic demonstration effects did not just make a fortune for individual entrepreneurs, but served to propel an entire technological system and to accelerate its dissemination world-wide. The first recurrent characteristic of long-wave behaviour is therefore directly connected to the second: the potential for very widespread application.

2.2. The recurrence of pervasive constellations of technical and organizational innovations

Each wave is characterized not just by one or two big innovations, nor even by a cluster of quite discrete individual innovations, but by a constellation of interdependent and mutually supportive technical and organizational innovations. As argued by Carlota Perez (1983), each of these constellations or paradigms has certain characteristics, which are common to them all. They all have identifiable and obvious *core inputs,* which have falling prices relative to other commodities during the critical transition period between one paradigm and the next. The principal producers and users of these inputs became the leading sectors (motive and carrier branches) in the upsurge of the economy. The demonstration effects occur relatively early in the diffusion of each new technological revolution and, whether they occur most conspicuously in firms making core inputs, in other leading sectors, or in associated infrastructures, they help to propel the diffusion of the whole constellation and not only a part of it.

It is not just the excitement generated by the first demonstration effects, important though these undoubtedly are, but the long-term potential which has become visible and which reverberates throughout the system as more and more applications of the new paradigm appear on the horizon. A second recurrent feature of the long waves is therefore that each one is characterized by the emergence and experimental testing of a new combination of inter-related innovations, which demonstrate remarkable gains in productivity and profitability at first in a few applications, but with the clear potential for very pervasive diffusion.

Ultimately, this full potential is realized in a period of prolonged prosperity but only after a structural crisis of adjustment, which can last. Examples of the pervasiveness of new technology systems in each new wave are the applications of steel and of electricity, of iron and steam power, of oil and internal combustion and, currently, of computers and new technologies of information and communication. The chip and the devices for communication are the key factors of the emerging long wave. As the following pages will indicate, these new devices are fundamental for creating new forms of economic production but also for generating new modes of cultural production.

2.3. The recurrence of waves of organizational and management changes in firms

A third recurrent feature of each revolution is that organizational and managerial changes introduced in the new leading sectors are widely imitated elsewhere. A new management style becomes fashionable and in the later waves in the twentieth century is disseminated by management consultants as well as through the media and social communication, propelled by example. The very success of the leading firms is sufficient in itself to stimulate imitative efforts in relation to their new management style but, of course, the technical innovations which they introduce are often also directly conducive to organizational changes in those firms which adopt them.

The use of computers and mobile communications are two obvious contemporary examples, but some organizational styles are not so directly dependent on technical innovations and have a momentum of their own. The sheer growth in the *size* of leading firms was itself an important factor in organizational and managerial changes in the nineteenth and twentieth centuries. The trends in organizational change are more complex than the narrowly technical changes but there is an identifiable recurrence of a new management style in each Kondratiev wave which influences many firms, although in diverse ways, throughout the economy.

This does not mean of course that *every* firm in *every* industry adopts a similar

“The very success of the leading firms is sufficient in itself to stimulate imitative efforts in relation to their new management style but, of course, the technical innovations which they introduce are often also directly conducive to organizational changes in those firms which adopt them”

management style or organizational structure. The idea of a *representative* firm characterising all firms is one which has been widely influential in economic theory, but it is not embraced here: on the contrary, evidence shows that with each technological revolution, the effects are very varied. With the mass production style, for example, firms in some industries were capable of introducing standardized products and using an assembly line resembling the Fordist line in the automobile industry. Many others continued to produce unique customized or small-batch products. Still others modified some features of the Fordist management style so that there were actually many varieties of Fordism, even within the automobile industry itself (Boyer, 1988). Only a minority of firms became recognizably 'Fordist'. Nevertheless, in industries as diverse as tourism, fast food, retail distribution and clothing, the influence of Fordist management philosophy and organizational change is clearly evident. Similarly with electrification, this led on the one hand to the growth of some giant electrical firms with specialized departmental management structures. On the other hand, it facilitated the de-centralized success of many small firms taking advantage of the new flexibility permitted by electric machinery, and the management of information-intensive processes allows for new forms of decentralization and delocalization.

2.4. Recurrent crises of structural adjustment

These examples show that there is some danger of making too schematic a model of the successive technological revolutions, which would do violence to their individual variety. This is especially the case because each one of them not only embodies a unique combination of products and processes but also affects other parts of the economy very unevenly, requiring different types of machinery, of materials and components, of distribution and of supporting services. Some entirely new branches of the economy make their appearance while other branches experience only marginal changes. Moreover, sometimes they affect particular *occupations* within industries and services which are otherwise little affected. The process of dissemination is therefore unpredictable and extremely uneven as new applications are explored, tested, expanded, modified or rejected. Nevertheless, a clearly observable and recurrent characteristic of each new technological revolution is its *pervasive* effect on the structure of the economic system. Although the *induced* branches of the economy are different, they are very significant in every case, and so too are the induced changes in skill requirements and hence in the education and training systems.

The fourth recurrent characteristic of each long wave is therefore a *crisis of structural adjustment* as the skills and distribution of the labour force and of firms adapt to the new paradigm, while the social conventions, contracts, laws and generally accepted procedures tend to change slowly and sometimes after periods of conflict.

Recurrent high levels of structural unemployment are an important manifestation of these adjustment crises in each long wave. The statistics for the nineteenth century are very poor, but there is strong evidence of very serious unemployment in the 1830s and 1840s in Britain, while David Wells (1890) commented on the widespread unemployment in most industrial countries in the 1880s and especially in those which were most advanced in the use of machinery. There is, of course, abundant statistical evidence of the heavy structural unemployment in the 1920s and 1930s and again in the 1980s and 1990s. Even in the 1920s boom in the United States, as Fearon (1987) and the NBER pointed out, there were sectors experiencing severe adjustment problems, such as coal, railways and ship-building. In Germany and Britain, heavy industry generally, but especially the steel industry and the ship-building industry experienced prolonged problems of structural adjustment. In the 1980s, the automobile industry, the oil industry, the synthetic materials industry and again the steel industry were among the many industries which experienced severe adjustment problems.

Nowadays, the crisis of structural adjustment expresses itself in a large army of permanently unemployed as a consequence of the mismatch of qualifications and employment between those sectors and branches with high profitability and competitiveness but few jobs, and those which have a large number of jobs but diminishing competitiveness. Without the reorganization of a virtuous match between the technological capacities and the social and institutional framework, the recovery will be slow and interrupted by deep crises, such as those triggered by the financial crashes of 1987 and 2007.

2.5. Recurrent changes in the regulatory regime

Finally, a recurrent feature of the qualitative changes engendered by the long wave is a periodic re-configuration of the *regime of regulation* of technology and of the economy more generally. It is quite obvious that such extensive changes as mechanization, electrification, motorization, and computerization raise entirely new requirements for education and training, which have led with each successive crisis of structural adjustment to a variety of movements for education reform. It is also obvious that each major new technology entails new requirements for safety and protection, whether of operatives in industry, consumers or people in certain exposed areas. However, the recurrent changes in the regulatory regime go well beyond these immediate and obvious induced effects. Even at this elementary level regulatory requirements can raise some fundamental *political* issues such as 'self-regulation' of industries versus state regulation, national versus international regulation or local versus national. They also raise questions of *standards* which tend also to become an area of conflict and dispute, both between competing groups seeking to promote their own version of the new technology, and between nations seeking to protect their own interests. Especially in the case of new infrastructural investment, questions of *ownership* and control also arise. If private ownership is the solution which is adopted in any particular case, this again immediately gives rise to questions of monopoly, competition and price regulation. Equally problematic are the questions of trade and protection, whether of new or of older industries.

Typically, the leaders in a new wave of technology, such as Britain in the nineteenth century or the United States in the twentieth century, will tend to advocate the opening up of world markets to the new products and services in which they excel, while catching-up countries will often deploy 'infant industry' arguments to justify various forms of protection. The leading economies will seek to advocate and, if they have the strength, to impose an international regulatory regime with institutions which promote the interests of their leading industries. Thus, what is at stake in each structural crisis is a re-constitution of the entire institutional and social framework because there is a mismatch between the regulatory framework developed and consolidated by a previous generation for older technologies and industries, and the needs of the newly emerging constellation and the interests of the new technological leaders.

Once a new technological and regulatory regime has become dominant and firmly established, the phenomenon of 'lock in' to the new regime becomes widely apparent. This is the case not only with lock in to dominant designs, technical standards, components and so forth, but also to all kinds of social standards and institutions, variable though these may be between different countries, in response to the changing balance of social and political forces in each country and on the international stage. The instability of the current international economic structure is revealed not by the challenge by emerging economies of the dominant role of the previous leaders, but instead by the fact that there is not a stable international order to make their trade coherent and to settle their disputes.

One of the aspects of this deep ongoing change, which is of interest for what follows, is the impact of economic and technological innovation on culture, altering previous modes of production of sense and images, distorting others, suggesting new ones, and creating universal references as part of a unified global market.

3. THE IMPACT OF ECONOMIC AND TECHNOLOGICAL INNOVATION ON THE PRODUCTION OF CULTURE

In order to exemplify this analysis of some historical trends, and to define their impact on the creation of values, I will turn now to the impact of innovation on the creation of cultural traits. Let me take the extreme example of art. It is extreme since art is posited and defines itself as autonomous from other social relations, and as a peculiar expression of creation of new culture or communication. The production of works of art as part of cultural production is that province where the autonomy in relation to social processes and, in particular, to economic determination, is more radically affirmed. Yet, art is an interesting case of the interaction of individuality and society, of technology conditioning invention.

Indeed, art has logic and a time of its own; quite often, it anticipates the future or constructs alternative worlds. That is a reason for proceeding in this section from the production of art to the general production of cultural artefacts, images and sounds or, in general, to references, as part of the cultural changes in a changing society.

In spite of the vindication of the autonomy of art, the producers live in concrete societies and their horizon is largely defined by the potentialities of their epoch. In the same sense, the creation of a specific culture, in the general sense of the coherence of forms of communication in fashion, food, literature, architecture, dance, or music, just as the evolution of languages or other social artefacts is largely bound by its particular epoch. The technological framework, the social structure and the historical process of the formation of knowledge define the setting for the work of art as for the construction of social cultures.

This section explores this connection, in order to present an overview of large processes of the social production of cultural products. This topic was not previously discussed in the literature, and it is only sketched here as a suggestion for further investigation. In this sense, I argue that there are specific tides in cultural production that can be better understood in the framework of the societal and historical vision indicated previously, and that this specific nexus is crucial for the understanding of the forms of organised social communication prevailing in modern societies.

I certainly do not suggest that technological means determine the production of cultural values as such; that would mean an underrating of the influence of social and individual traits in the creation of cultural artefacts and communication. But evidence shows how the landscape of technological opportunities shaped different forms of cultural production, amplifying and selecting new means and inducing new trajectories such as the cinema, the video, the clip, and the continuous production of images and messages.

In a nutshell, the argument is, first, that major changes in the economy are causally related to dramatic alterations in the social structure, including in the forms of work and power, as well as the dominant modes of communication, and, second, that the trends in cultural production are unintelligible outside the context of these changes.

“The argument is, first, that major changes in the economy are causally related to dramatic alterations in the social structure, including in the forms of work and power, as well as the dominant modes of communication, and, second, that the trends in cultural production are unintelligible outside the context of these changes”

The next table takes up this argument. Each epoch is defined by the industrial revolution originating the *maelstrom* that changed the way of life and shaped each specific experience of modernity. Consequently, three main categories are indicated: while the original industrial revolution set the tone for the process of *modernization*, following on from the slow development after the Enlightenment and the sixteenth century, *modernism* was the emerging language for the conflict against the enlargement and dominance of the modern market. Not against technology or machines—indeed they were worshipped by the futurists and other modernists, who praised the automobile as the archetype of the progress of humankind—but against the impersonal relations of the market and the general reification under capitalism. Finally, the triumph of the market over its radical opposition was marked by its extension into what was, until then, the partially separated world of art production: this period has been called *postmodernism*. As Jameson puts it, “modernism [was] the experience and the result of *incomplete* modernization, (…) [and] the postmodern begins to make its appearance whenever the modernization process no longer has archaic features and obstacles to overcome and has triumphantly implanted its own autonomous logic" (Jameson, 1991: 366).

That said, Modernism yesterday was not, and Postmodernism today is not either culturally dominant or even absolutely hegemonic: as indicated in the table, they can be considered as emerging traits in cultural production, representative of the ongoing conflict. But Modernism and Postmodernism are the trends most closely associated with the fractures of history in their own time. They were indeed seen by contemporaries as earthquakes—a celebrated example is that of the much quoted and tragic images of *Angelus Novus* by Paul Klee, a representation of the transformation imposed by the second industrial revolution and the motif for Walter Benjamin's much quoted reminiscences:

> An angel looking as though he is about to move away from something he is fixedly contemplating. His eyes are staring, his mouth is open, his wings are spread. This is how one pictures the angel of history. His face is turned toward the past. Where we perceive a chain of events, he sees one single catastrophe which keeps piling wreckage and hurls it in front of his feet. The angel would like to stay, awaken the dead, and make whole what has been smashed. But a storm is blowing from Paradise: it has got caught in his wings with such violence that the angel can no longer close them. This storm irresistibly propels him into the future to which his back is turned, while the pile of debris before him grows skyward. This storm is what we call progress. (Benjamin, 1973: 259-260)

This wave of progress invaded daily life and transformed the modes of production, distribution and communication at the end of the nineteenth century. The culture of the new century was part of this catastrophe: "Il faut être absolument moderne", said Rimbaud. The next industrial and technological revolution imposed a new version of the dictum: we cannot but be postmodern. The structure of this evolution is the theme for the table below:

Table 1. The production of culture in a long-term perspective

Period	Industrial revolution	Emerging cultural traits	Examples	Modes of communication	Centres of cultural production
Long Wave II 1848-1890s	Machine production of steam-driven engines	**Realism**		Conversation Books	London
Transition 1890s					Paris
Long Wave III 1890s-1940s	Machine production of electric and internal-combustion engines	**Modernism**	Expressionism, Cubism, Futurism	Printed word (books, newspapers)	Paris, Berlin, Vienna, St. Petersburg
Transition 1940s-50s			Rock'n Roll	Radio, Film	New York
Long Wave IV 1940s-...	Machine production of electrical and electronic machines	**Post Modernism**	Pop, Punk, Fashion, Advertisement	Film, TV	New York Los Angeles Bollywood
Transition ?			MTV, Youtube, Facebook, Twitter	TV, Electronically transmitted bits, clips	

As the table indicates, there are considerable lags between causally connected events and trends; moreover, there is a large margin of autonomy between technological transformations, allowing for new methods of dissemination and permitting new experiences of the process of modernization, and their cultural counterparts. Yet the creation of new technological means sets the pace for the transformation.

The undisputed but not unique example is the creation of the 'Gutenberg Galaxy' in the fifteenth century: it allowed for the development 'of a system essentially dominated by the typographic mind and the phonetic alphabet order' (Castells, 1996, I: 331). As the alphabet was the dominant 'conceptual technology' since 700 BC Greece, it established itself as the privileged infrastructure for the codification of cumulative knowledge. But it became a dominant mode of communication just when the industrial capacity established the printed word as the direct form of expression and the book as the cultural tool of the elite. Consequently, for the whole period sounds and images were outside the scope of written discourse and were relegated to the domain of the separate and slightly esoteric artistic production.

A new epoch was then opened when Fordist production spread to the whole social fabric and extended to the mechanical reproduction of works of art. Radio and film—the first distinctively mediatic art form, since *opera* had only performed that function in restricted areas of Europe, such as Italy, as had been the case with theatre in Britain—then became the dominant modes of communication.

Finally, and we come to our own time, a 'new alphabetic order', a new digital meta-language is being imposed as the cultural infrastructure: 'A technological transformation of similar historical dimensions [as the creation of the alphabet] is taking place 2700 years later, namely the integration of various modes of communication into an interactive network' (Castells, ibid.: 328). These epoch-making transformations are the theme for the next table.

As previously indicated, the epochs depicted in Tables 1 and 2 were or are not periods of uniformity, and their cultural

Table 2. Modes of cultural production

Period	Dominant character-istic	Construction of meanings	Technique of cultural production	Socially constructed attitude	Emerging trend in cultural production
LW I and II	Liberal capitalism	Hierarchical dissemination of knowledge, auratic works	Discontinuous and scarce production of individual works	Reverence Admiration	Promethaic modernization
LW III	Fordism	Mechanical reproduction of the works of art, distantiation effect	Discontinuous artefacts and dense production	Contemplation Concentration	Apollonian modernism
LW IV, emerging V	Late capitalism	Anti-auratic, populist culture, global reification effect	Continuous flow, networks	Distraction	Dyonisian postmodernism

production was *a fortiori* a turbulent landscape of diversity and contradiction. Their emerging cultural *traits* were not necessarily dominant or hegemonic throughout the period, although they marked a peculiar vision of the storm of modernization, and eventually constituted the more distinguishable features of its inheritance.

Realism represented the first interpretation of the changing world, and this new world reserved a specific role for the entertainment business: popular novels, theatre, and popular opera were, for some countries, the forerunners of the cultural industry of the second half of the century. Although this business was still separated from the production of high culture, the dissemination of the market in this direction anticipated the aestheticization of daily life—but even this would still require another major technological change.

Modernism was the response to these first moves: breaking with the aesthetics of representation in art and the theoretical discourse based on the insulated worlds of culture and social life (Lash and Urry, 1987: 13), the modernist revolution was built on challenges of the capitalist process of modernization. It led to non-representational and expressionist painting, to new lyric poetry, to existentialism in philosophy, to the *films d'auteur*. Attacking the market and not technology, the modernists were fascists with Marinetti and communists with Maiakovsky: they praised cars, speed and movement, neat colours and strong feelings. Picabia, Duchamp, Fernand Léger, Diego Rivera painted machines and the possibility of reconstruction of social life in a new technological world; Frank Lloyd Wright, Le Corbusier and Mies applied the new vision to architecture and rebuilt urban life.

But modernism was rooted in high culture, and the headquarters of resistance were established in the authenticity, originality and uniqueness of the work of art: its discourse was that of creativity (Lash and Urry, 1987: 286) and the defence of the aura of artistic artefacts. In that sense, art in society vindicated a radical separation of cultural forms from the social framework—and this was the reason for its intense appeal as well as for its failure. In a matter of decades, the expansion of the market conquered this last bastion of cultural critique and transformed it into an industry, and in particular into an industry of production in continuous flows.

In fact, the crucial alteration introduced in the post-war period was the widespread diffusion of commercial TV. Consequently, the film industry, the epicentre of cultural production since the beginning of the century, was transmuted from a production of episodic and unique pieces, seen by large audiences in unique settings, into a production of flows of images and sounds to be seen simultaneously in all private settings. The simultaneous collective experience was transmuted into a simultaneous individualised experience. Furthermore, the modifications went deeper than the setting of the experience of the cultural product: they imposed a change in the product itself, since the continuous flow abolished the effort of memory and imposed the loss of historicity, mixing news, films, soaps and contests at the same level of discourse. All sounds and images were reduced to bits of *infotainment*. The great consequence was the fully used potential for the construction of 'fictive temporalities' and therefore the 'technological appropriation of subjectivity', generating a specific and novel type of media populism that was to become the basis for the entertainment industry (Jameson, 1991: 74).

The social consequences of this dramatic change in culture are still to be fully understood. In any case, they were the result of the transformation of culture by the market. The growing importance of advertising, the consumption of the discourse of consumption and the narrative of desire inscribed in publicity constitute the image as the final form of reification of the commodity: the product is identified with its brand or logo. Advertising is the dominant form of production of signs in postmodern culture: postmodernism is that specific mode of production in which advertisement is the new technology of communication, a new 'conceptual technology' or 'alphabetic order' of our days. As a consequence, culture has a function for the market.

Fashion and fast food, B-films and remakes, Warhol's pop art, parodies and kitsch, science fiction, music and video reduced to clips, these images populate this universe of *pastiche*, to use Thomas Mann's concept. Categories of space have replaced categories of time, historical depth has been lost to ephemerity and concentration has been replaced by superficial trivia: commodification of culture is a devastating process.

As this process is our contemporary, its implications are still largely unknown, although there are two that follow from the pattern of communication imposed by this cultural revolution. First, 'a crucial effect of the electronic media and spatio-temporal changes in our disorganizing capitalist societies has been the decentring of identities and the loosening or destructuration of group and grid' (Lash and Urry, 1987: 299). But, second, this iconography of modernity also imposes a bipolar opposition between the Net and the Self, so that 'in this condition of structural schizophrenia between function and meaning, patterns of social communication become increasingly under stress' (Castells, 1996, I: 3).

“Advertising is the dominant form of production of signs in postmodern culture: postmodernism is that specific mode of production in which advertisement is the new technology of communication, a new 'conceptual technology' or 'alphabetic order' of our days. As a consequence, culture has a function for the market”

How did machines then produce machines and information produce information to the extent that we are transported from a culture of virtual reality to a culture of real virtuality in this period of transition? This is the question for Castells, in his magnificent *The Information Age: Economy, Society and Culture*. The answer resides in the technological changes associated with the information and communication revolution emerging through the fourth long wave, and in the concrete process of social selection that has determined the shape of the new techno-economic paradigm challenging the still prevailing mode of development in our times of mismatch and transition. The answer is the network of cultural products and facilities of communication in a market economy.

In this framework, we follow Jameson's suggestion to reconsider the concept of 'late capitalism' as it was used by the Frankfurt school, namely by Adorno and Horkheimer, and lately by Ernest Mandel. *Late capitalism* describes the galaxy of economic structures, methods of production and cultural *substrata* derived from the expansion of commodification towards Nature and the Self or the Unconscious. This is a process of reification of all social relations, i.e. one purer form of capitalism. Late capitalism is thus the name for the technological transformations diffused since the 1950s and for the cultural alterations emerging from the 1960s until the present. As a cultural constellation, it had a long period of maturation: it was even anticipated in the early decades of the century by Dada and Surrealism, which invented these postmodernist tones, although they rooted their activity in a mood of denunciation of the market economy as the adversary to art. Yet it was when the technology became available for the production of continuous flows of infotainment that postmodernism won the day.

Contrary to MacLuhan and so many other commentators, this victory did not represent the imposition of a complete universal culture: we do not live in a global village, but in 'customized cottages globally produced and locally distributed' (Castells, 1996-I: 341). Each cultural artefact is locally bounded and the production of icons is still mediated by national and regional frontiers: their understanding is largely local. The global world is a world of diversity. But icons are industrially produced and are the constitutive bits and clips of our social communication, and this aesthetic of distraction is universal. It marks the triumph of a new conceptual technology on the map of culture.

Yet, this technology does not by itself impose a social order; on the contrary, its prevalence depends on the social mutations here described as the long waves of capitalist development.

4. CONCLUSION: SOCIAL AND CULTURAL CHANGES IN THE LONG WAVES

The preliminary presentation of recurrent changes characteristic of each long wave has already gone beyond purely economic and technological phenomena, and the previous section discussed the production of cultural references, which is largely autonomous although influenced by the social movement as a whole. As the crisis of structural adjustment and the periodic changes in the regulatory regime raise fundamental questions of the relationship between technical change, political change and cultural change, this is shortly evoked as follows:

First, consider changes in the regulatory regime, whether at a national or international level, since they can lead to the most fundamental political and ideological conflicts within and between nations. Lloyd-Jones and Lewis (1998) have made a particularly valuable study of the conflict over the Corn Laws in the 1830s and 1840s in Britain and the later conflict on Tariff Reform in Britain in the late nineteenth and early twentieth centuries. Both of these conflicts split the ruling Tory Party from top to bottom and led to major re-alignments in British politics and each was associated with a long-wave structural crisis. In the same sense, the problems of tariff protection also had profound effects in the United

States, Germany and Japan as they were industrialising and catching up in technology. However, the political dimensions of free trade and tariff reform clearly go far beyond just the question of regulating some new products and services, or protecting older industries, even though these problems may trigger the conflicts. Fundamental national interests, as well as those of particular industries, are often felt to be at stake and friction over trade issues can be a major source of friction in international relations more generally, as illustrated in the Anglo-German naval armaments race before 1914. The World War between 1939-1945, marking the turning point after the depressive long wave of the first decades of the century and opening an epoch of growth and prosperity, is another illustration of this concatenation of political and military solutions for long-term disputes over markets and resources.

Second, consider the depth of the social clashes which may be exacerbated during a structural crisis, which is illustrated no less clearly by the labour conflicts that are engendered. History registers the widespread social unrest as well as the outbreaks of 'Luddism' associated with the destruction of old crafts and occupations, such as those of the hand-loom weavers. Some historians have argued that Luddism, especially in the hosiery industry in Nottinghamshire was inspired mainly by the desire to protect British quality standards in foreign trade. The workers supposedly feared more for the loss of jobs through the erosion of British sales in foreign markets than simply from mechanization. Whatever the interpretation may be, it is fairly obvious that the destruction of the livelihood of hundreds of thousands of people is bound to be a cause of acute social unrest and this has indeed been the case in every crisis of structural adjustment. There are also bound to be conflicts within the expanding industries and technologies over pay, status and working conditions for various groups of managers and workers. Modern conflicts raise a wider range of problems, with deep cultural implications, such as concerns about the ecological sustainability of industrial or urban policies, and the effects on climate change, international relations and poverty issues.

The third domain of interest here is the technical changes, which are relatively unrelated to other changes previously described. This is only superficially known, since it is widely accepted that the evolution, for instance, of the ship, of the hammer, of flints for tools and weapons, of the harnessing of the horse, and of the steam engine or the plough emphasize alike the *relative* autonomy of the improvements which were made over the centuries to these artefacts, so essential for human civilization. The selection environment, which interests, inspires and constrains engineers, designers, inventors and mechanics and many historians of technology, is primarily the *technical* environment, the criteria of technical efficiency and reliability and of compatibility with existing or future conceivable technology systems.

The reciprocal influence of science and technology upon each other has been demonstrated in numerous studies and is indeed obvious in such fields as computer technology and biotechnology today, as well as in earlier developments such as thermodynamics and the steam engine. Technology has to take account of the laws of nature and hence of science. Nevertheless, Derek Price (1984), Rosenberg (1969, 1982), Pavitt (1995) and many others have produced cogent arguments for recognizing the special features of each sub-system

> “The reciprocal influence of science and technology upon each other has been demonstrated in numerous studies and is indeed obvious in such fields as computer technology and biotechnology today, as well as in earlier developments such as thermodynamics and the steam engine”

precisely in order to understand the nature of their interaction. Nor does this refer only to recent history, as the massive contribution of Needham (1954) to the history of Chinese science and technology clearly illustrate.

Historians of technology, such as Gille (1978) and Hughes (1982), have amply demonstrated the *systemic* nature of technologies and analyzed the interdependencies between different elements in technology systems. Both they and Rosenberg have also shown that the technological imperatives derived from these systemic features may serve as focusing devices for new inventive efforts. Finally, in their seminal paper ‘In Search of Useful Theory of Innovation’ Nelson and Winter (1977) drew attention to the role of *technological trajectories*, both specific to particular products or industries and general trajectories such as electrification or mechanization affecting a vast number of processes and industries, including compur. They rightly identified the combination of such trajectories with scaling-up in production and markets as one of the most powerful influences on economic growth. These ideas were further developed by Dosi (1982) in his work on technological trajectories and technological paradigms, in which he pointed to the relative autonomy of some patterns of technological development by analogy with Kuhn’s paradigms in science. Despite the obvious close interdependence between technology and the economy or technology and science, it is essential to take into account these relatively autonomous features in the history of technology.

A satisfactory theory of economic growth and development must certainly take account of these processes, but it should also recognize that the *relative* autonomy of evolutionary developments in science and technology justifies some independent consideration. An essentially similar argument applies to *economic* change. No-one can seriously doubt the importance of capital accumulation, profits, changes in company organization and the behaviour of firms and of banks for the evolution of industrial societies over the past two centuries. Economic institutions too have some relative autonomy in the cycles of their development. In any case, explanations of economic growth must pay especially close attention to the interdependencies between economic history and technological history. It is precisely the need to understand the changing nature of this interdependency which leads to the proposal of a theory of recurrent phenomena and ‘out of sync’ phases of development, when, for example, changes in technology may outstrip the institutional forms of the production and market system that may be slow to change or impervious to change for relatively long periods. The reverse may also occur, providing impetus to new technological developments, as with the assembly line or factory production.

Finally, *cultural* change is generally accepted as an important influence on economic growth. In the previous section I explored the opposite sense of influence: that from the economic mode of development and availability of technologies to the production of culture, the former creating opportunities and incentives for new developments of the latter. It is in order to emphasise now the impact of culture, as part of the creation of social values, on the dynamics of growth, since these values tend to concentrate the resistance or suspicion about characteristics of the institutional change imposed by the diffusion of the clusters of radical innovations. Social values are shaped by institutions and recognise contracts, laws, routines, types of communication, hierarchies, the forms of each social pact ruling each society, and tend therefore to be adverse to radical and unknown change. Although some societies—some cultures—are more inclined to accept the challenge of innovation and rupture in the previous trajectory, it is understandable that a flexible answer to the hurricane of change is to impose rules that are previously known to the society. Indeed, any new economy or new technology is appropriated according to the previously established knowledge. This is why modern developed societies are so stable: they change but tend to adapt to change. Evolutionary economics is certainly familiar with these processes, since they mimic natural evolution so well, with the creation of variety (i.e. innovation) and the selection of change (i.e. stable environment).

But this process of adaptation and creation of stability is also responsible for some conservatism against radical implications of social innovation flowing from radical innovation in the techno-economic system, where they tend to originate. The socio-institutional system and its cultural standards tend to generate the mismatch or desynchronicity previously indicated as the engine of the long waves of capitalist development.

In any case, a general view of the cultural determinations of social relations should emphasize all these contributions to the formation of social mentalities and modes of reasoning, including the motivation for accepting change and routine. Indeed, social, political and cultural changes interact in modern societies under the impact of technical and organizational changes, either to react or to resist. If retardation or acceleration phenomena are to be explained, the dominant cultures of an epoch reveal and register the combined effects of its histories of present and past. Institutions, which are the result of such cultures and social relations, are the decisive structures for economic evolution and the condition for growing or perishing.

The long waves, this modern curse of Heraclitus, encapsulate the social and economic processes of evolution and change, of adaptation and of creation of variety, or innovation. But innovation alters structures and culture, institutions and routines, which are locked in established trajectories. This is why the epoch-making radical innovations, clustering in industrial and technological revolutions, can create new upsurges in economic development but tend to confront large institutional resistance.

This is the case today. The early years of the 21st century were marked both by the magnificent extension of a cluster of innovations applicable to a wide range of productive processes and service economy, and simultaneously by the disarrangement of the financial markets, with speculation destroying wealth and savings, and accumulation undermining the creation of value. This is explainable by the asynchronous movements in the depressive phase of the long wave and by the emergence of new profitable branches, stimulating speculation and over-accumulation. As in previous long waves, the crucial question is not why the technological push does not translate into macroeconomic performance, but instead, how should the social networks be reassembled and economic institutions created that are able to regulate a new system, creating jobs, qualifications, welfare and further innovation. Some will say this is finally an issue of changing culture.

BIBLIOGRAPHY

Benjamin, W. (1973), *Illuminations*, London: Fontana.
Berman, M. (1983), *All That is Solid Melts into Air – the Experience of Modernity*, London: Verso.
Boyer, R. (1988), "Technical Change and the Theory of Regulation", in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds.), *Technical Change and Economic Theory*, London: Pinter, pp. 67-94.
Castells, M. (1996), *The Information Age: Economy, Society and Culture* I, *The Rise of the Network Society*, Oxford: Blackwell.
Dosi, G. (1982), "Technological Paradigms and Technological Trajectories", *Research Policy*, 11(3), pp. 147-62.
Fearon, P. (1987), *War, Prosperity and Depression: The American Economy 1917-1945*, Deddington: Phillip Alan.
Freeman, C., F. Louçã (2002), *As Time Goes By – From the Industrial Revolutions to the Information Revolution*, Oxford: Oxford University Press.
Gille, B. (1978), *Histoire des Techniques*, Paris: Gallimard.
Goodwin, R. (1985), 'A Personal Perspective Economics', *Banca Nazionale del Lavoro Quarterly Review*, pp. 152, 3-13.
Harvey, D. (1992), *A Condição Pós-Moderna*, São Paulo: Loyola.
Hughes, T. (1982), *Networks of Power Electrification in Western Society, 1800-1930*, Baltimore: Johns Hopkins University Press.
Jameson, F. (1991), *Postmodernism or the Cultural Logic of Late Capitalism*, London: Verso.
Lash, S., and J. Urry, (1987), *The End of Organized Capitalism*, Madison: University of Wisconsin Press.
Lloyd-Jones, R., and M. Lewis (1998), *British Industrial Capitalism since the Industrial Revolution*, London: UCL Press.
Mandel, E. (1980), *Long Waves of Capitalist Development*, Cambridge: CUP.
Nelson, R., and S. Winter (1977), "In Search of a Useful Theory of Innovation", *Research Policy*, 6(1), pp. 36-76.
Pavitt, K. (1995), "Academic Research and Technical Change", in MacLeod, R. (ed.), *Technology and the Human Prospect*, London: Pinter, pp. 31-54.
Poletayev, A. (1987), "Profits and Long Waves", working paper, Montpellier Conference on Long Waves.
Price, D. (1984), "The Science-Technology Relationship", *Research Policy*, 13(1), pp. 3-20.
Rosenberg, N. (1969), "Directions of Technological Change: Inducement Mechanisms and Focusing Devices", *Economic Development and Cultural Change* 18, pp. 1-24.
Rosenberg, N. (1982), *Inside the Black Box: Technology and Economics*, Cambridge: CUP.
Sombart, W. (1929), "Economic Theory and Economic History", *Economic History Review* 2: p. 18.
Tylecote, A. (1992), *The Long Waves in the World Economy: the Present Crisis in Historical Perspective*, London: Routledge.
Wells, D. (1890), *Recent Economic Changes*, London: Longman.

BBVA

Technological change and the evolution of the U.S. "National Innovation System", 1880-1990

David C. Mowery
Haas School of Business, University of California, Berkeley & NBER

INTRODUCTION

Technological change has played a central role in US economic growth since the 19th century. The pioneering work of Solow (1957) and Abramovitz (1956) both suggested that expansion in labor and capital accounted for no more than 15% of total growth in US output per head between the middle of the 19th century and the 1950s. The remaining 85%, labeled the "residual," is widely interpreted as a measure of the economic effects of technological change, although Abramovitz famously referred to it as a "measure of our ignorance". This essay explores the changing characteristics of innovation and the relationship between innovation and US economic growth during this lengthy period.

The transition from the 19th to the 20th centuries was accompanied by a shift in the sources of US economic growth from exploitation of a rich domestic endowment of natural resources to the exploitation of "created" resources based on knowledge and trained scientists and engineers. Advances in technology and knowledge aided the exploitation of the US resource endowment during the 19th century, enabling the United States to overtake the global economic leader of the time, Great Britain. Beginning in the late 19th century, however, the United States embarked on a prolonged transition from resource-led to knowledge-led economic growth.

Institutional innovation was an indispensable complement to technological innovation during and after this period in US economic development. Public and private investments in new organizational structures for the support of knowledge creation, innovation and education were essential to the changing trajectory of US economic growth in the 19th and 20th centuries. State and federal government investments supported the creation of a higher education infrastructure that eventually proved to be an important source of scientific and engineering knowledge and personnel (Goldin and Katz, 2009). Industrial investment in the development of new technologies also made important contributions during the 20th century. And the 1945-89 period, dominated by geopolitical tensions that sparked massive investment of public funds in defense and related missions by the federal government, witnessed a further transformation of this

> "The pioneers in this organizational innovation were the large German chemicals firms of the last quarter of the 19th century, whose growth was based on innovations in dyestuffs"

complex mix of public and private institutions devoted to supporting innovation.

This essay surveys the development of the US "national innovation system" from the late 19th to the late 20th centuries. The "national innovation system" framework for analyzing innovative performance and policy is the subject of a substantial body of scholarship that has flourished since the first articulation of the concept in Freeman (1987; see also Lundvall, 1992 and Nelson, 1993). "National" innovation systems typically include the institutions, policies, actors, and processes that affect the creation of knowledge, the innovation processes that translate research into applications (either for commercial sale or deployment in such "nonmarket" contexts as national defense), and the processes that influence the adoption of innovations.

Accordingly, the US national innovation system includes not just the institutions performing R&D and the level and sources of funding for such R&D, but policies—such as antitrust policy, intellectual property rights, and regulatory policy—that affect technology development, the training of scientists and engineers, and technology adoption. Institutional elements, such as national systems of higher education and corporate finance and governance, represent other important components of national innovation systems. The structure of a nation's innovation system is the result of complex historical processes of institutional development that are affected by public policy and other influences. Moreover, the performance of these systems depends in part on the actions and decisions of private enterprises that can reinforce or offset the effects of government policies.

OVERVIEW OF US ECONOMIC "CATCH-UP," 1800–1910

US economic growth during the 19th century has been characterized by Abramovitz and David (2000), David and Wright (1997), and Wright (2007) as more capital- and natural resource-intensive than Western European growth during the same period. The capital-intensive trajectory of US economic growth during the 19th century reflected the high rates of investment and significant innovation in the transportation and communications infrastructure (canals, railways, the telegraph and telephone) that contributed to the development of another major factor in 19th-century US economic growth—the large, unified domestic market that manufacturers in particular exploited in the wake of the Civil War. Through much of the 19th century, this domestic market was characterized by relatively low levels of income inequality, by comparison with Great Britain and other European economies, resulting in a large, homogeneous profile of consumer demand. Reliable all-weather inland transportation also facilitated the export of the produce of the abundant and relatively fertile expanse of farmland within the United States.

During the last two decades of the 19th century, the US economy began a prolonged transition from the extensive growth trajectory that relied on expanding capital, resource, and labor inputs to a more knowledge-intensive growth trajectory that was associated with higher rates of total factor productivity growth (Abramovitz and David, 2002). One of the most dramatic illustrations of this gradual shift was the increased exploitation of scientific and technical knowledge in US resource extraction industries that began in the late 19th century (David and Wright, 1997). As David and Wright pointed out, the United States pioneered in the development of new institutions for research and education in mining engineering, geology and related fields that supported expansion in US output of minerals and related raw materials during this period. Based in part on a growing endowment of economically relevant natural resources, US firms had moved to the technological frontier in mass-production manufacturing, particularly in metalworking and machinery industries, by the late 19th century (Nelson and Wright, 1994: 135).[1]

Many of the first academic institutions specializing in these fields of research and education were publicly funded, illustrating another important characteristic of the post-1870 period of economic catch-up. The 1862 Morrill Act established a foundation for publicly funded higher education, and (along with the 1887 Hatch Act) expanded federal and state government funding for research and extension activities in agriculture. The development of mass higher education in the United States occurred in parallel with the emergence of the first US "research university" (Johns Hopkins University, founded in 1876), which was based on the German research university structure that had proven to be effective in supporting scientific research and collaboration with industry. Although decades (and billions in public funding) were required to bring US universities to positions of global scientific leadership, even before their attainment of research excellence these institutions played a crucial role in training generations of scientists, engineers, and managers, and developed networks of collaboration in scientific and technical research with US industry that contributed to US economic growth in the late 19th and 20th centuries.

Much of the technological innovation that drove US economic development during the 19th century was "pre-scientific," relying as much on trial-and-error experimentation by skilled practitioners as on activities that might be described as "R&D". The reliance of 19th century innovation on "tinkering" declined in the final decades of the century, with the growth of new areas of industrial production and innovation that relied on more complex technologies that were linked to the frontiers of scientific and engineering knowledge. Their reliance on more formalized knowledge meant that the growth of the "new industries" of the Second Industrial Revolution, particularly chemicals and electrical machinery, was associated with investments in R&D within the firm, an activity with little precedent in most US firms.

The pioneers in this organizational innovation were the large German chemicals firms of the last quarter of the 19th century, whose growth was based on innovations in dyestuffs. But by the early 20th century, a number of large US firms had also established in-house R&D organizations.

[1] "These new turn-of-the-century achievements may be thought of as the confluence of two technological streams: the ongoing advance of mechanical and metalworking skills and performance, focused on the high-volume production of standardized commodities; and the process of exploring, developing, and utilizing the mineral resource base of the national economy.

The growth of these laboratories almost certainly could not have occurred without complementary changes in institutions external to the firm, ranging from the development of US universities to the growth of new mechanisms for industrial finance. Nonetheless, the rise of the industrial R&D laboratory represented a fundamental shift in the structure of the US national innovation system.

THE GROWTH OF US INDUSTRIAL RESEARCH IN THE "SECOND INDUSTRIAL REVOLUTION," 1890-1945

By the first decade of the 20th century, a number of large US manufacturing firms had established in-house industrial research laboratories as part of a broader restructuring that transformed their scale, management structures, product lines, and global reach. Many of the earliest US corporate investors in industrial R&D, such as General Electric and Alcoa, were founded on product or process innovations that drew on advances in physics and chemistry. The corporate R&D laboratory brought more of the process of developing and improving industrial technology into the boundaries of US manufacturing firms, reducing the technological and economic importance of the independent inventor (Schmookler, 1957).

But the in-house research facilities of large US firms were not concerned exclusively with the creation of new technologies. Like the laboratories of the German dyestuff firms, these US industrial laboratories also monitored technological developments outside the firm and advised corporate managers on the acquisition of externally developed technologies. Many of Du Pont's major product and process innovations during this period, for example, were obtained from sources external to the firm, and Du Pont further developed and commercialized them (Mueller, 1962; Hounshell and Smith, 1988; Hounshell, 1995).[2] In-house R&D in US firms developed in parallel with independent R&D laboratories that performed research on a contract basis (see also Mowery, 1983a). But over the course of the 20th century, contract-research firms' share of industrial research employment declined.

The evolution of industrial research in the United States was influenced by another factor that was absent in Germany during the late 19th and early 20th centuries -- competition policy. By the late 19th century, judicial interpretations of the Sherman Antitrust Act had made agreements among firms for the control of prices and output targets of civil prosecution. The 1895-1904 merger wave in the United States, particularly the surge in mergers after 1898, was one response to this new legal environment. Since informal and formal price-fixing and market-sharing agreements had been declared illegal in a growing number of cases, firms resorted to horizontal mergers to control prices and markets.[3]

The Sherman Act's encouragement of horizontal mergers ended with the Supreme Court's 1904 Northern Securities decision, but many large US firms responded to the new antitrust environment by pursuing strategies of diversification that relied on in-house R&D to support the commercialization of new technologies that were developed internally or purchased from external sources. George Eastman saw industrial research as a means of supporting the diversification and growth of Eastman Kodak (Sturchio, 1988, p. 8). The Du Pont Company used industrial research to diversify out of the black and smokeless powder businesses even before the 1913 antitrust decision that forced the divestiture of much of the firm's

[2] The research facilities of AT&T were instrumental in the procurement of the "triode" from independent inventor Lee de Forest, and advised senior corporate management on their decision to obtain loading-coil technology from Pupin (Reich, 1985). General Electric's research operations monitored foreign technological advances in lamp filaments and the inventive activities of outside firms or individuals, and pursued patent rights to innovations developed all over the world (Reich, 1985: 61). The Standard Oil Company of New Jersey established its Development Department precisely to carry out development of technologies obtained from other sources, rather than for original research (Gibb and Knowlton, 1956: 525). Alcoa's R&D operations also closely monitored and frequently purchased process innovations from external sources (Graham and Pruitt, 1990: 145-147).

[3] See Stigler (1968). The Supreme Court ruled in the Trans Missouri Association case in 1898 and the Addyston Pipe case in 1899 that the Sherman Act outlawed all agreements among firms on prices or market sharing. Data in Thorelli (1954) and Lamoreaux (1985) indicate an increase in merger activity between the 1895-1898 and 1899-1902 periods. Lamoreaux (1985) argues that other factors, including the increasing capital-intensity of production technologies and the resulting rise in fixed costs, were more important influences on the US merger wave, but her account (p. 109) also acknowledges the importance of the Sherman Act in the peak of the merger wave. Lamoreaux also emphasizes the incentives created by tighter Sherman Act enforcement after 1904 for firms to pursue alternatives to merger or cartelization as strategies for attaining or preserving market power.

black powder and dynamite businesses (Hounshell and Smith, 1988: 57).

Although it discouraged horizontal mergers among large firms in the same lines of business, US antitrust policy through much of the pre-1940 period had little effect on efforts by these firms to acquire new technologies from external sources. The development of industrial research, as well as the creation of a market for the acquisition and sale of industrial technologies, also benefited from reforms in US patent policy between 1890 and 1910 that strengthened patent-holder rights (See Mowery, 1995).[4] Judicial tolerance for restrictive patent licensing policies further increased the value of patents in corporate research strategies. Although the search for new patents provided an incentive to pursue industrial research, the impending expiration of these patents created another important impetus for the establishment of industrial research laboratories. Both American Telephone and Telegraph and General Electric, for example, established or expanded their in-house laboratories in response to the intensified competitive pressure that resulted from the expiration of key patents (Reich, 1985; Millard, 1990: 156). Intensive efforts to improve and protect corporate technological assets complemented the acquisition of patents in related technologies from other firms and independent inventors.

Many of the elements of the "Open Innovation" model, defined by its leading proponent as a new model for managing corporate innovation in which "firms can and should use external ideas as well as internal ideas" (Chesbrough, 2003), were present in the early development of US industrial R&D. The in-house R&D facilities of leading industrial firms served as monitors of external technological developments that supported the purchase by their parent firms of important innovations from independent inventors and other firms.

Another area in which the pre-1940 era in the development of industrial research resembles that of the past two decades is the evidence of collaborative linkages between industrial and academic research. Furman and MacGarvie (2005) show that pharmaceuticals industry R&D facilities founded during 1927– 46 in the United States tended to locate near leading research universities, and provide other evidence of university-industry collaboration in pharmaceuticals during this period. Other scholars (Mowery *et al.,* 2004; Rosenberg, 1998) have emphasized the importance of university-industry collaboration during this period, not least in the development of such important fields of university research as chemical engineering.

Training by public universities of scientists and engineers for employment in industrial research also linked US universities and industry during the first decades of the 20th century. The Ph.D.s trained in public universities were important participants in the expansion of industrial research employment during this period (Thackray, 1982: 211).[5] The size of this trained manpower pool was as important as its quality; although the situation was improving in the decade before 1940, Cohen (1976) noted that virtually all "serious" US scientists completed their studies at European universities. Thackray *et al.* (1985) argue that American chemistry research during this period attracted attention (in the form of citations in other scientific papers) as much for its quantity as its quality.

Federal expenditure for R&D throughout the 1930s constituted 12-20% of total US

[4] These technology-acquisition strategies built on a domestic market for intellectual property that grew substantially during the 1880-1920 period. According to Lamoreaux and Sokoloff (1999), the development of a national market for intellectual property enabled independent inventors to specialize and thereby enhanced their productivity and the overall innovative performance of the US economy during this period. By the early 20th century, however, the increased costs of inventive activity and greater demand for formal scientific and engineering training led to the supplanting of independent by corporate inventors (Lamoreaux and Sokoloff, 2005).

[5] Hounshell and Smith (1988: 298) report that 46 of the 176 Ph.D.s overseen by Carl Marvel, longtime professor in the University of Illinois chemistry department, went to work for one firm, Du Pont. According to Thackray (1982: 221), 65% of the 184 Ph.D.s overseen by Professor Roger Adams of the University of Illinois during 1918-58 went directly into industrial employment. In 1940, 30 of the 46 Ph.D.s produced by the University of Illinois chemistry department were first employed in industry.

R&D expenditure, and industry accounted for about two-thirds of the total. The remainder came from universities, state governments, private foundations, and research institutes. One estimate suggests that state funds may have accounted for as much as 14% of university research funding during 1935-36 (National Resources Planning Board, 1942: 178). Moreover, the contribution of state governments to non-agricultural university research appears from these data to have exceeded the federal contribution, in sharp contrast to the postwar period. The modest role of the federal government in financing US R&D during the 1930s changed radically as a result of the political events of the next 20 years.

THE TRANSFORMATION OF THE US INNOVATION SYSTEM, 1945-1989

The global conflict of 1939-1945 transformed the structure of R&D throughout the industrial economies. In few if any other industrial economies, however, was this transformation as dramatic as in the United States. The structure of the pre-1940 US R&D system resembled those of other leading industrial economies of the era, such as the United Kingdom, Germany, and France: industry was a significant funder and performer of R&D, and central government funding of R&D was modest. But the postwar US R&D system differed from those of other industrial economies in at least three aspects: 1. US antitrust policy during the postwar period was unusually stringent; 2. small, new firms played an important role in the commercialization of new technologies, especially in information technology;[6] and 3. defense-related R&D funding and procurement exercised a pervasive influence in the high-technology sectors of the US economy.

A central characteristic of the institutional transformation of the US national innovation system during this period was increased federal support for R&D, most of which was defense-related. Defense-related R&D spending accounted for more than 80% of total federal R&D spending for much of the 1950s, and rarely has dropped below 50% of federal R&D expenditure during the entire 1949-2005 period (figure 1; data from US Office of Management and Budget, 2005). Since federal R&D spending accounted for more than 50% of total national R&D spending during 1953-78 (data for overall national R&D investment are available only after 1952), and only dropped below 40% in 1991 (its postwar low point of 25% appeared in 2000, as Figure 2 shows; data from National Science Board, 2006), the significance of the federal government's defense-related R&D investment is obvious—in some years during the postwar period (e.g., the late 1950s and early 1960s), public defense-related R&D investment accounted for nearly one-half of total national R&D spending.

Defense-related R&D programs affected innovation throughout the postwar US economy. Much of the "R&D infrastructure" of the postwar economy, including large research facilities in industry, government, or academia, was built with funding from defense-related R&D programs. In addition, defense-related funding for academic research in fields ranging from computer science to oceanography supported the training of thousands of scientists and engineers. A second important channel of influence was associated with technological "spinoffs" -- technological advances developed for defense-related applications that found large markets in the civilian economy. Such spinoffs proved to be

[6] Chandler and Hikino (1997) argue that established firms dominated the commercialization of new technologies in most sectors of the postwar US economy, with the significant exception of "...electronic data-processing technologies, based on the transistor and integrated circuit..." (p. 33).

Figure 1. Federal and Nonfederally funded R&D, 1953-2002

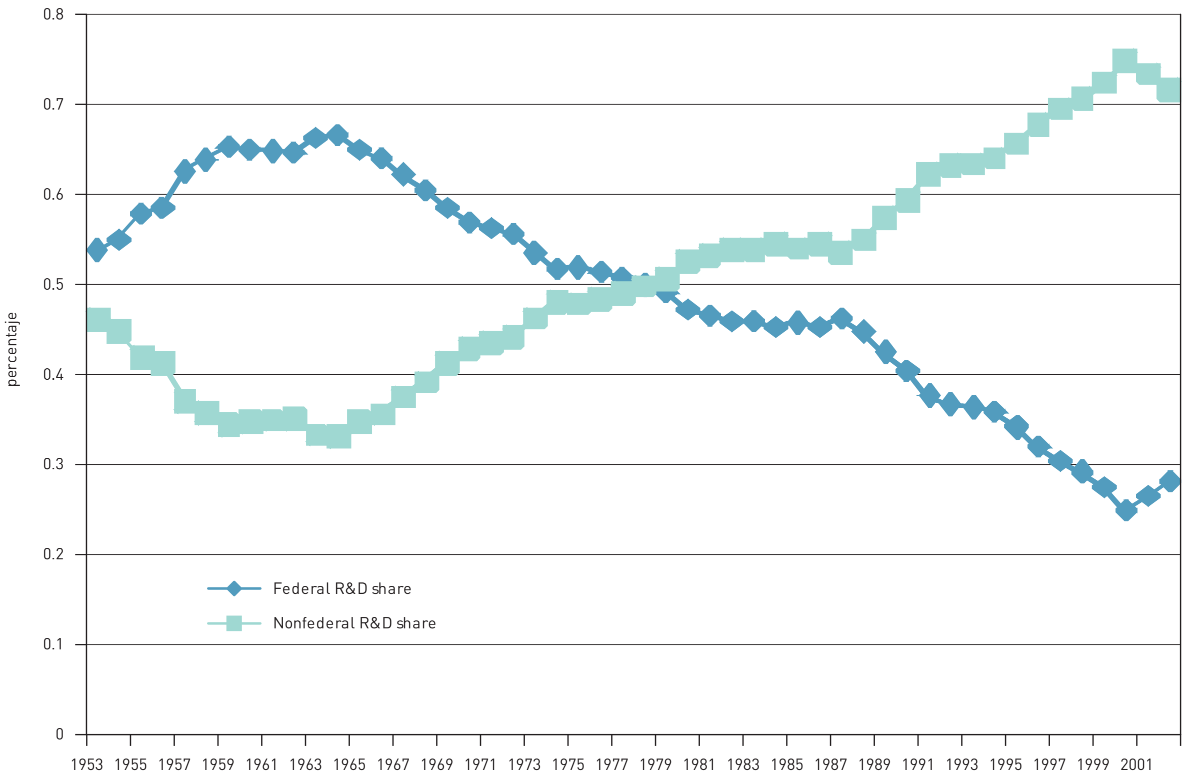
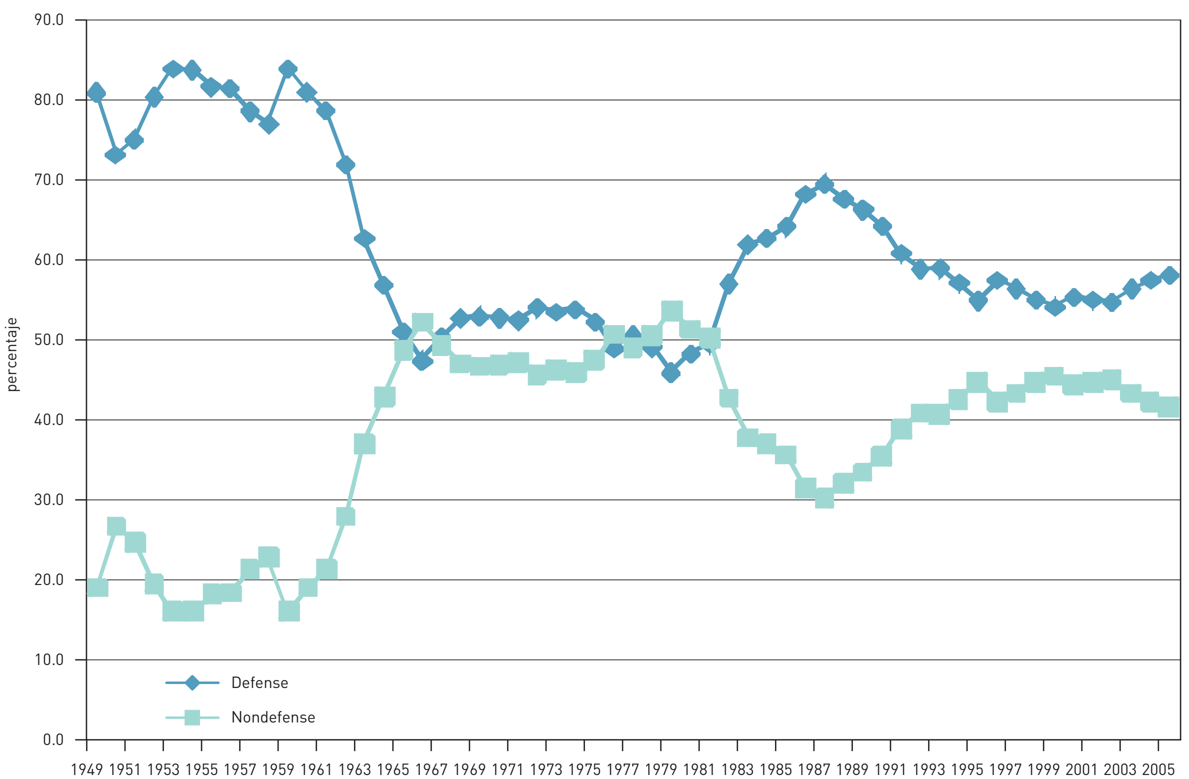


Figure 2. Defense & Nondefense share of total federal gov't R&D outlays, 1949-2005



particularly significant in aerospace and information technology.

A third important channel through which defense-related spending on new technologies advanced civilian technological applications, aiding the exploitation of technological "spinoffs", was procurement. Postwar defense-related R&D programs typically were complemented by substantial purchases of new technologies. The US military services, whose procurement requirements typically emphasized performance above all other characteristics (including cost), played a particularly important role during the post-1945 period as a "lead purchaser," placing large orders for early versions of new technologies. These procurement orders enabled suppliers of products such as transistors or integrated circuits to reduce the prices of their products and improve their reliability and functionality.[7] Government procurement allowed innovators to benefit from production-related learning and cost reductions by expanding output of early versions of a new technology. Reductions in production costs led to lower prices for the technologies, by opening up civilian markets, which typically are more price-sensitive.

Examples of technological "spinoffs" from defense-related R&D spending in the postwar United States include the jet engine and swept-wing airframe that transformed the postwar US commercial aircraft industry. Major advances in computer networking and computer memory technologies, which found rapid applications in civilian as well as military programs, also trace their origins to defense-supported R&D programs. Defense-related procurement was particularly important in the postwar US information technology industry. In other areas, however, such as numerically controlled machine tools, defense-related demand for applications of novel technologies had detrimental effects on the commercial fortunes of US machine tool firms (Mazzoleni, 1999; Stowsky, 1992). And the light-water nuclear reactor technologies that were first developed for military applications proved to be poorly adapted to the civilian sector (Cowan, 1990).

The "spinoff" and "procurement" channels of interaction were most significant when defense and civilian requirements for new technologies overlapped significantly and/or when defense-related demand accounted for a large share of total demand for a new technology. In both aerospace and information technology, the economic and technological significance of military-civilian spinoffs appear to have declined as a result of growing divergence in the technological requirements of military and civilian products, as well as the growth of civilian markets for these products. Moreover, in some cases, such as information technology, the influence of defense applications on the overall direction of technical development not only declined by the 1990s; defense technologies in some areas lagged behind those in the civilian sector, reflecting the reduced influence of defense-related demand and R&D investment on the innovative activities of private firms.

Although defense-related R&D programs typically are dominated by spending on "development," rather than "research, the sheer size of the overall investment of public funds meant that government defense-related R&D supported academic research in a diverse array of disciplines in the physical sciences and engineering. But federal R&D funding in the bio-medical sciences, which was allocated largely to research, also grew substantially during the post-1945 period. Although the primary federal funder of

[7] New technologies undergo a prolonged period of "debugging", performance and reliability improvement, cost reduction, and learning on the part of users and producers about applications and maintenance (Mowery and Rosenberg, 1999). The pace and pattern of such progressive improvement affect the rate of adoption, and the rate of adoption in turn affects the development of these innovations.

> "Major advances in computer networking and computer memory technologies, which found rapid applications in civilian as well as military programs, also trace their origins to defense-supported R&D programs"

biomedical research, the National Institute of Health (NIH), was established in 1930, its extramural research program received significant support only after the founding in 1937 of the National Cancer Institute, the first of 28 research institutes within the NIH (Swain, 1962) and during the late 1940s, NIH's extramural research programs began to grow more rapidly.[8] By 1970, NIH funding of academic research amounted to $2 billion (in 2000 dollars), which had grown to more than $13 billion by 2009.

Rapid growth in the NIH budget, along with slower growth in defense-related R&D after 1970, shifted the disciplinary composition of federally funded research away from the physical sciences and engineering and toward biomedical research. Growth in federally funded biomedical R&D has been more than matched by growth in privately funded R&D investment in the US pharmaceuticals industry since 1990. By the early 21st century, federally funded R&D spending accounted for less than 40% of overall R&D spending in this sector.[9] The NIH now supports half of all federal non-defense R&D and over 60% of federally funded research in American universities.[10]

NIH support of academic research contributed to the scientific advances in molecular biology and related fields that gave rise to the biotechnology industry during the 1970s and 1980s. Scientific advances at such universities as Columbia, Stanford, and the University of California at San Francisco held out considerable potential for applications in pharmaceuticals and related industries. All three of these universities, as well as others, became important "incubators" for new firms, and increasingly patented faculty discoveries. Even before the passage of the Bayh-Dole Act of 1980, important patents had been filed on behalf of these three universities, and university licensing in biomedical fields grew rapidly during the 1980s and 1990s (See Mowery *et al.,* 2004).

In contrast to federal investments in IT, federal R&D policy in the biomedical sector did not combine federal procurement-related "demand-pull" with its large investments in research. But the dominance of third-party payment (from both public and private sources) for the majority of US healthcare meant that patients and doctors alike were more responsive to performance than to price. As a result, new technologies tended to command a higher price premium in the United States biomedical market than in other industrial economies, where public insurance systems often limited prices and margins. These incentives to adopt and apply new technologies quickly may well have influenced the commercial exploitation by US pharmaceutical, medical device, and biotechnology firms of the knowledge and techniques produced by NIH R&D investments.

[8] A substantial majority (80%) of the annual research budget of the NIH supports research conducted in laboratories at universities, generally in medical schools.

[9] The US Pharmaceutical Manufacturers Association estimated that foreign and US pharmaceuticals firms invested more than $26 billion in R&D in the United States in 2002, substantially above the $16 billion R&D investment by the National Institute of Health in the same year (See Pharmaceutical Manufacturers Association, 2003, for both estimates).

[10] National Science Foundation/ Division of Science Resources Statistics, Survey of Research and Development Expenditures at Universities and Colleges, FY 2006. http://www.nsf.gov/statistics/nsf08300/pdf/nsf08300.pdf

As I noted earlier, an internationally unique characteristic of the US national innovation system that dates back to the late 19^{th} century has been the unusually stringent character of US antitrust policy, which exerted great influence on the early R&D strategies of many leading US industrial firms. Antitrust policy continued to affect the development of industrial R&D during the postwar period. US antitrust policy during the 1950s and 1960s made it more difficult for large US firms to acquire firms in "related" technologies or industries, and increased their reliance on intra-firm sources for new technologies (see Fligstein, 1990). In the case of Du Pont, the use of the central laboratory and Development Department to seek technologies from external sources was ruled out by senior management as a result of perceived antitrust restrictions on acquisitions in related industries. As a result, internal discovery (rather than development) of new products became paramount (Hounshell and Smith, 1988 emphasize the firm's postwar expansion in R&D and its search for "new nylons"[11]), in contrast to the firm's R&D strategy before World War II. The inward focus of Du Pont research appears to have impaired the firm's postwar innovative performance, even as its central corporate research laboratory gained a sterling reputation within the global scientific community.

In other US firms, senior managers sought to maintain growth through the acquisition of firms in unrelated lines of business, creating conglomerate firms with few if any technological links among products and processes. Chandler (1990) and others (e.g., Ravenscraft and Scherer, 1987; Fligstein, 1990) argue that diversification weakened senior management understanding of and commitment to the development of the technologies that historically had been essential to the competitive success, eroding the quality and consistency of decision-making on technology-related issues.[12]

Another novel characteristic of the US national innovation system during the 1945-90 period -- one that contrasted with the pre-1940 period --was the prominence of new firms in commercializing new technologies. In industries that effectively did not exist before 1940, such as computers, semiconductors, and biotechnology, new firms played important roles in the commercialization of innovations. These postwar US industries differed from their counterparts in Japan and most Western European economies, where established electronics and pharmaceuticals firms retained dominant roles in the commercialization of these technologies.

Several factors contributed to the importance of new firms in the postwar US innovation system. The large basic research establishments in universities, government, and a number of private firms served as "incubators" for the development of innovations that "walked out the door" with individuals who established firms to commercialize them. Although Klepper (2009) argues that a similar pattern of entrepreneurial exit and establishment of new firms within the same geographic region also characterized the US automobile industry in the early 20^{th} century, the evolution of the postwar US biotechnology, microelectronics and computer industries was heavily affected by such new-firm "spinoffs" from established firms. Indeed, high levels of labor mobility within regional agglomerations of high-technology firms served as an important channel for technology diffusion and as a magnet for other firms in related industries to locate in these areas. Such labor mobility also aided

[11] Hounshell and Smith (1988) and Mueller (1962) both argue that discovery and development of nylon, one of Du Pont's most commercially successful innovations, was in fact atypical of the firm's pre-1940 R&D strategy, which bore more than a passing resemblance to "open innovation." Rather than being developed to the point of commercialization following its acquisition by Du Pont, nylon was based on the basic research of Carothers within Du Pont's central corporate research facilities. The successful development of nylon from basic research through to commercialization nevertheless exerted a strong influence on Du Pont's postwar R&D strategy, not least because of the fact that many senior Du Pont executives had direct experience with the nylon project. Hounshell (1992) argues that Du Pont had far less success in employing the "lessons of nylon" to manage such costly postwar synthetic fiber innovations as Delrin.

[12] Graham's discussion (1986) of the failure of RCA to commercialize its videodisk technology in the face of the firm's extensive diversification into such unrelated industries as automobile rental agencies and frozen food is an illustrative analysis of the failures of technology management that accompanied the conglomerate-diversification strategies of many US firms in the 1960s and 1970s.

in the transfer of knowledge and know-how within many of these nascent high-technology industries.[13] The importance of new firms in commercializing postwar innovations in these new industries in the postwar US economy also relied on the extension to much smaller firms of the equity-based system of industrial finance that distinguished the US economy from those of Germany and Japan.

CONCLUSION

Along with other industrial economies, the United States shifted from an economy whose performance was based on the exploitation of domestic natural resources, including agricultural resources, to a "knowledge-based economy" during the 20th century. This transition took decades, but it also was characterized by a number of phenomena widely cited as hallmarks of 21st-century innovation. "Open innovation," for example, in which large corporations utilize intra-firm capabilities to scan the technological horizon for potential acquisitions of new technologies, accurately describes the strategies of many of the large US corporate pioneers of in-house R&D during the early 20th century. Their acquisitions of technologies from external sources also relied on the operation of a market for intellectual property that was widespread during the early decades of the 20th century, although its importance was subsequently supplanted by the in-house technology development activities of large firms.

This brief survey also highlights the close interaction among technological, policy, and institutional influences within the evolution of the US national innovation system. The discussion underscores the linkages between the processes of technological innovation and adoption that are essential to economic growth in all industrial economies. Much of the economic influence of post-1945 federal R&D spending, for example, flowed from the effects of public policy on both support for the development of new technologies and support for their rapid adoption. Moreover, in fields such as information technology, the widespread adoption by US users of such innovations as desktop computers and computer networking created a vast domestic platform that supported user-led innovation. For this "general purpose technology" in particular, innovation and adoption interacted and accelerated one another. Public policies to address future technological challenges such as global climate change or public health must take into account the importance of consistency and support for both technological innovation and adoption.

[13] Discussing the development of laser technology, Bromberg (1991) highlights the importance of linkages among research funders and performers within the United States during the 1950s and 1960s that in turn were based on researcher mobility: "Academic scientists were linked to industrial scientists through the consultancies that university professors held in large and small firms, through the industrial sponsorship of university fellowships, and through the placement of university graduates and postdoctoral fellows in industry. They were linked by joint projects, of which a major example here is the Townes-Schawlow paper of [sic] optical masers, and through sabbaticals that academics took in industry and industrial scientists took in universities. Academic scientists were linked with the Department of Defense R&D groups, and with other government agencies through tours of duty in research organizations such as the Institute for Defense Analyses, through work at DoD-funded laboratories such as the Columbia Radiation Laboratory or the MIT Research laboratory for Electronics, and through government study groups and consultancies. They were also linked by the fact that so much of their research was supported by the Department of Defense and NASA." (Bromberg, 1991: 224).

REFERENCES

ABRAMOVITZ, M. (1956), "Resource and Output Trends in the United States since 1870", *American Economic Review* 46, pp. 5-23.

ABRAMOVITZ, M., and P. A. DAVID (2000), "American Macroeconomic Growth in the Era of Knowledge-Based Progress", in S. L. Engerman and R. E. Gallman (eds.), *The Cambridge Economic History of the United States*, vol. III., New York: Cambridge University Press.

American Association for the Advancement of Science (2005), *Research and Development in the FY2006 Budget,* Washington, D.C.: American Association for the Advancement of Science.

BOK, D. (1982), *Beyond the Ivory Tower,* Cambridge, MA: Harvard University Press.

BRAUN, E., and S. MACDONALD (1978), *Revolution in Miniature,* New York: Cambridge University Press.

BROMBERG, L. (1991), *The Laser in America, 1950-1970*, Cambridge, MA: The MIT Press.

BUSH, V. (1945), *Science: The Endless Frontier*, Washington, D.C.: U.S. Government Printing Office.

CHANDLER, A. D., Jr. (1977), *The Visible Hand*, Cambridge, MA: Harvard University Press.

CHANDLER, A. D., Jr. (1990), *Scale and Scope,* Cambridge: MA: Harvard University Press.

CHANDLER, A. D., Jr. (2001), *Inventing the Electronic Century*, New York: Free Press.

CHANDLER, A. D., Jr., and T. HIKINO (1997), "The large industrial enterprise and the dynamics of modern economic growth", in A. D. Chandler, Jr., F. Amatori and T. Hikino (eds.), *Big Business and the Wealth of Nations*, Cambridge, UK: Cambridge University Press.

CHESBROUGH, H. (2003), *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Boston, MA: Harvard Business School Press.

CHESBROUGH, H. (2006), "Open Innovation: A New Paradigm for Understanding Industrial Innovation" in H. Chesbrough, W. Vanhaverbeke and J. West (eds.), *Open Innovation: Researching a New Paradigm*, New York: Oxford University Press.

COHEN, I. B. (1976), "Science and the Growth of the American Republic", *Review of Politics,* 38, pp. 359-398.

COHEN, W., R. FLORIDA and R. GOE (1994), "University-Industry Research Centers in the United States", Technical Report, Center for Economic Development, Carnegie-Mellon University.

COHEN, W., R. FLORIDA, L. RANDAZZESE and J. WALSH (1998), "Industry and the Academy: Uneasy Partners in the Cause of Technological Advance", in R. Noll, ed., *Challenges to the Research University*, Washington, D.C.: Brookings Institution.

COWAN, R. (1990), "Nuclear Power Reactors: A Study in Technological Lock-In", *Journal of Economic History* 50, pp. 541-567.

DAVID, P. A., and G. WRIGHT (1997), "Increasing Returns and the Genesis of American Resource Abundance", *Industrial and Corporate Change,* 6, pp. 203-245.

DAVIS, L. A., and R. J. CULL (2000), "International Capital Movements, Domestic Capital Markets, and American Economic Growth, 1820-1914", in S. L. Engerman and R. E. Gallman (eds.), *The Cambridge Economic History of the United States*, vol. II, New York: Cambridge University Press.

EDQUIST, C. (2004), "Systems of Innovation: Perspectives and Challenges", in J. Fagerberg, D. C. Mowery and R. R. Nelson (eds.), *Oxford Handbook of Innovation and Policy*, Oxford, UK: Oxford University Press.

FLIGSTEIN, N. (1990), *The Transformation of Corporate Control,* Cambridge, MA: Harvard University Press.

FREEMAN, C. (1987), *Technology Policy and Economic Performance: Lessons from Japan*, London: Pinter.

FREEMAN, C. (1995), "The 'National System of Innovation' in Historical Perspective", *Cambridge Journal of Economics,* 19, pp. 5-24.

FURMAN, J., and M. MACGARVIE (2005), "Early Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry", National Bureau of Economic Research working paper #11470.

GALAMBOS, L., and J. L. STURCHIO (1998), "Pharmaceutical firms and the Transition to Biotechnology: A Study in Strategic Innovation", *Business History Review* 72, pp. 250-278.

GALLMAN, R. E. (2000), "Economic Growth and Structural Change in the Long Nineteenth Century", in S. L. Engerman and R. E. Gallman (eds.), *The Cambridge Economic History of the United States*, vol. II, New York: Cambridge University Press.

GERSCHENKRON, A. (1962), *Economic Backwardness in Historical Perspective* Cambridge, MA: Harvard University Press.

GIBB, G.S., and E. H. KNOWLTON (1956), *The Resurgent Years: History of Standard Oil Company (New Jersey), 1911-1927*, New York: Harper & Row.

GOLDIN, and KATZ (2009), *The Race Between Education and Technology*, Cambridge, MA: Harvard University Press.

GRAHAM, M. B. W. (1986), *RCA and the Videodisc: The Business of Research,* New York: Cambridge University PressGraham, M. B. W., and H. G. Pruitt (1990), *R&D for Industry: A Century of Technical Innovation at Alcoa*, New York: Cambridge University Press.

HAINES, M. R. (2000), "The Population of the United States, 1790-1920" in S. L. Engerman and R. E. Gallman (eds.), *The Cambridge Economic History of the United States*, vol. II, New York: Cambridge University Press.

HOUNSHELL, D. A. (1995), "Du Pont and the Management of Large-Scale Research and Development", in P. Gallison and B. Hevly, *Big Science: The Growth of Large-Scale Research*, Stanford, CA: Stanford University Press.

HOUNSHELL, D. A. (2000), "The Medium is the Message, or How Context Matters: The RAND Corporation Builds an Economics of Innovation, 1946-62", in T. P. Hughes and A. Hughes (eds.), *Systems, Experts, and Computers,* Cambridge, MA: MIT Press.

HOUNSHELL, D. A., and J. K. SMITH (1998), *Science and Corporate Strategy: Du Pont R&D, 1902-1980,* New York: Cambridge University Press.

KLEPPER, S. (2009), "Silicon Valley—A Chip off the old Detroit Bloc" in Z. Acs, B. Audretsch and R. Strom (eds.), *Entrepreneurship, Growth, and Public Policy,* New York: Cambridge University Press.

LAMOREAUX, N. R., and K. L. SOKOLOFF (1999), ìInventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuriesî, in N. R. Lamoreaux, D. M. G. Raff and P. Temin (eds.), *Learning by Doing in Markets, Firms, and Countries*, Chicago: University of Chicago Press.

LAMOREAUX, N. R., and K. L. SOKOLOFF (2005), "The Decline of the Independent Inventor: A Schumpeterian Story", National Bureau of Economic Research Working Paper #11654.

LANGLOIS, R. (2003), "The Vanishing Hand: The Changing Dynamics of Industrial Capitalism", *Industrial and Corporate Change*, pp. 12, 351-385.

LANGLOIS, R. N., and D. C. MOWERY (1996), "The Federal Government Role in the Development of the U.S. Software Industry" in D. C. Mowery (ed.), *The International Computer Software Industry: A Comparative Study of Industry Evolution and Structure*, New York: Oxford University Press.

LESLIE, S. W. (1993), *The Cold War and American Science,* New York: Columbia University Press.

LEVIN, Richard C. (1982), "The Semiconductor Industry", in R. R. Nelson ed., *Government and Technical Progress: A Cross-Industry Analysis,* New York: Pergamon Press.

LUNDVALL, B. -Å. (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London: Pinter.

MADDISON, A. (1994), "Explaining the Economic Performance of Nations, 1820-1989", in W. J. Baumol, R. R. Nelson and E. N. Wolff (eds.), *Convergence of Productivity*, New York: Oxford University Press.

MARR, W.L., and D. G. PATERSON (1980), *Canada: An Economic History*, Toronto: Macmillan.

MAZZOLENI, R. (1999), "Innovation in the Machine Tool Industry: A Historical Perspective on the Dynamics of Comparative Advantage", in D. C. Mowery and R. R. Nelson (eds.), *The Sources of Industrial Leadership,* New York: Cambridge University Press.

MCMILLAN, G. S., F. NARIN and D. S. DEEDS (2000), "An Analysis of the Critical Role of Public Science in Innovation: The Case of Biotechnology", *Research Policy* 29, pp.1-8.

MILLARD, A. (1990), *Edison and the Business of Invention,* Baltimore, MD: Johns Hopkins University Press.

MOWERY, D. C.(1983a), "Industrial Research, Firm Size, Growth, and Survival, 1921-1946", *Journal of Economic History,* 43, pp. 953-980.

Mowery, D. C. (1983b), "The Relationship between the Contractual and In-House Forms of Industrial Research in American Manufacturing, 1900-1940", *Explorations in Economic History,* 20, pp. 351-374.

Mowery, D. C. (1995), "The Boundaries of the U.S. Firm in R&D", in N. R. Lamoreaux and D. M. G. Raff (eds.), *Coordination and Information: Historical Perspectives on the Organization of Enterprise*, Chicago: University of Chicago Press for NBER.

Mowery, D. C. (2005), "National Security and National Innovation Systems", presented at the PRIME/PREST workshop on "Reevaluating the Role of Defence and Security R&D in the Innovation System", University of Manchester, September 19-21.

Mowery, D. C., R. R. Nelson, B. Sampat and A. Ziedonis (2004), *Ivory Tower and Industrial Innovation,* Stanford, CA: Stanford University Press.

Mowery, D. C., and N. Rosenberg (1999), *Paths of Innovation*, New York: Cambridge University Press.

Mowery, D. C., and B. N. Sampat (2004), "The Bayh-Dole Act of 1980 and University-Industry Technology Transfer: A Model for Other OECD Governments?" *Journal of Technology Transfer,* 30, pp. 115-127.

Mowery, D. C., and T. Simcoe (2002), "The History and Evolution of the Internet" in B. Steil, R. Nelson and D. Victor (eds.), *Technological Innovation and Economic Performance,* Princeton, NJ: Princeton University Press.

Mowery, D. C., and A. Ziedonis (2002), "Academic Patent Quality and Quantity Before and After the Bayh-Dole Act in the United States", *Research Policy,* 31, pp. 399-418.

Mueller, W. F. (1962), "The Origins of the Basic Inventions Underlying DuPont's Major Product and Process Innovations, 1920 to 1950" in *The Rate and Direction of Inventive Activity*, Princeton: Princeton University Press.

National Research Council (1999), *Funding a Revolution: Government Support for Computing Research*, National Academies Press.

National Research Council (2004), *A Patent System for the 21st Century,* Washington, D.C.: National Academies Press.

National Resources Planning Board, *Research—A National Resource*, Washington, D.C.: U.S. Government Printing Office.

National Science Board (2006), *Science and Engineering Indicators: 2006*, Washington, D.C.: U.S. Government Printing Office.

National Science Foundation, Directorate for Social, Behavioral, and Economic Sciences (1995a), "1993 Spending Falls for U.S. Industrial R&D, Nonmanufacturing Share Increases", http://www.nsf.gov/statistics/databrf/sdb95325.pdf.

National Science Foundation, Science Resources Division (1995b), "Research and Development in Industry, 1992", http://www.nsf.gov/statistics/nsf96333/appa.pdf.

National Science Foundation, Science Resources Division (2005), *Academic Research and Development Expenditures: Fiscal Year 2003*, Washington D.C.: National Science Foundation.

Neal, A. D., and D. G. Goyder (1980), *The Antitrust Laws of the United States*, Cambridge, UK: Cambridge University Press.

Nelson, R. R. (ed.) (1993), *National Innovation Systems: A Comparative Analysis*, New York: Oxford University Press.

Nelson, R. R., and G. Wright (1992), "The Rise and Fall of American Technological Leadership", *Journal* of Economic Literature 30(4), December, pp. 1931-1964.

Patit, J. M., S. P. Raj, and D. Wilemon (2006), "Integrating Internal and External R&D: What Can We Learn from the History of Industrial R&D?", presented at the meetings of the Academy of Management, Atlanta, GA, August 4-6.

Pharmaceutical Manufacturers Association (2003), *Industry Profile 2003,* http://www.phrma.org/publications/profile02.

Ravenscraft, D., and F. M. Scherer (1987), *Mergers, Sell-Offs, and Economic Efficiency,* Washington, D.C.: Brookings Institution.

Reich, L. S. (1985), *The Making of American Industrial Research*, New York: Cambridge University Press.

Rosenberg, N. (1998), "Technological Change in Chemicals: The Role of University-Industry Relations" in A. Arora, R. Landau, and N. Rosenberg, eds., *Chemicals and Long-Term Economic Growth*, New York: John Wiley.

Schmookler, J. (1957), "Inventors Past and Present", *Review of Economics and Statistics,* 39, pp. 321-333.

Schumpeter, J. A. (1943), *Capitalism, Socialism, and Democracy,* New York: Harper & Row.

Solow, R. M. (1957), "Technical Change and the Aggregate Production Function", *Review of Economics and Statistics*, 39, pp. 312-320.

Stigler, G. J. (1968), "Monopoly and Oligopoly by Merger", in G. J. Stigler (ed.), *The Organization of Industry,* Homewood, IL, Irwin.

Stowsky, J. (1992), "From Spin-off to Spin-on: Redefining the Military's Role in American Technology Development", in W. Sandholtz, M. Borrus, J. Zysman, K. Conca, J. Stowsky, S. Vogel, and S. Weber, *The Highest Stakes,* New York: Oxford University Press.

Sturchio, J. L. (1988), "Experimenting with Research: Kenneth Mees, Eastman Kodak, and the Challenges of Diversification", presented at "The R&D Pioneers", Hagley Museum and Library, October 7.

Swann, J. P. (1988), *Academic Scientists and the Pharmaceutical Industry,* Baltimore, MD, Johns Hopkins University Press.

Thackray, A. (1982), "University-Industry Connections and Chemical Research: An Historical Perspective", in *University-Industry Research Relationships*, Washington, D.C., National Science Board.

Thackray, A., J. L. Sturchio, P.T. Carroll and R. Bud (1985), *Chemistry in America, 1876-1976: Historical Indicators*, Dordrecht: Reidel.

Thorelli, H.B. (1954), *Federal Antitrust Policy,* Baltimore, MD: Johns Hopkins University Press.

Trow, M. (1979), "Aspects of Diversity in American Higher Education", in H. Gans, *On the Making of Americans*, Philadelphia: University of Pennsylvania Press.

Trow, M. (1991), "American Higher Education: 'Exceptional' or Just Different?" in B. E. Shafer (ed.), *Is America Different? A New Look at American Exceptionalism,* New York: Oxford University Press.

U.S. Office of Management and Budget (2005), *Budget of the U.S. Government, Fiscal Year 2006: Historical Tables*, Washington, D.C.: U.S. Government Printing Office.

Wildes, K.L., and N. A. Lindgren (1985), *A Century of Electrical Engineering and Computer Science at MIT, 1882-1982* Cambridge, MA: MIT Press.

Wright, G. (2007), "Historical Origins of the New American Economy", unpublished MS, Stanford University.

BBVA

Mapping Creativity in the European Union:

The role of labour market structure and systems of education and training[1]

Edward Lorenz
Université de Nice Sophia Antipolis

Bengt-Åake Lundvall
Aalborg Universitet

1. INTRODUCTION

Creativity has attracted the attention of researchers in a variety of disciplines including behavioural psychology and management. Within the field of psychology the focus has been primarily on the relationship between creativity and such individual attributes as intelligence, knowledge and personality (Barron and Harrington, 1981; Helson, 1996; Sternberg, 1988; Sternberg and Lubart, 1991; Weisberg, 1993).[2] In the management literature the focus has been more on how creativity emerges from the interaction between the individual employee and various aspects of management style and work organisation. Woodman, Sawyer, and Griffin (1993), for example, see creativity as resulting from the interaction of individual, group and organizational variables. Amabile *et al.* (1996) similarly focus on social and organizational factors, arguing in particular that creativity at work is supported by organizational and supervisory encouragement as well as by a diversity of ideas within the work group (Bharadwag and Menon, 2000; Drazin, Glynn and Kazanjian, 1999; and Ford, 1996).

Although there has been some work on the cultural or systemic basis of creativity (Csikszentmihalyi, 1988; Lubart, 1999), prior to Richard Florida's publication of *The Rise of the Creative Class* (2002), relatively little attention has been given to analysing the basis for and the impact of creativity at the levels of regions and nations. By putting forward creativity as the driving force of economic growth, and by presenting the rise of creativity as a general account of the current transformation of the economy comparable to the knowledge-based economy hypothesis, Florida's research has done more than any of the more specialised research to bring creativity to the forefront of debate in the social sciences. Further, in a series of empirical studies focusing on the relationship between investments in human capital, creativity and regional economic performance, Florida and his co-authors have argued that the creative class provides a new and alternative standard to the level of educational attainment for measuring human capital in studies focusing on regional development (Florida, Mellander and Stolarick, 2008; Mellander and Florida, 2006).

[1] This chapter draws on the analysis in an article that we originally published in the *Cambridge Journal of Economics*, 2010.

[2] See R. J. Sternberg (ed.), 1999, for an overview of the literature.

Drawing inspiration from Florida's research as well as from the more specialised research on creativity in the fields of behavioural psychology and management, this paper sets out firstly to provide what we believe to be the first mapping of creativity at work for the 27 member nations of the European Union. We show that there are significant national differences in the likelihood of employees being involved in creative work activity even after adjusting for national differences in occupational and industrial structure.

Referring to recent research that extends the perspective on national systems of innovation to include the organisation of work, labour market characteristics and education systems (Hall and Soskice 2001; Lam and Lundvall, 2006; Lorenz and Lundvall 2006; Lundvall, 2002; Whitley 2006), we then explore by means of aggregate correlations at the national level the relationship between creativity at work and characteristics of national labour markets and systems of education and training. Among the key findings presented in this paper is that the level of creative work activity tends to be higher in nations with broad competence-based systems of education and training that place value on equality of access to life-long learning opportunities and the continuing acquisition of job-related skills. The level of creativity also tends to be higher in nations characterised by the combination of high levels of labour market mobility and well-developed systems of unemployment protection (flexible security) and with active labour market policies.

The chapter is structured as follows. Section 2 develops the mapping of the importance of creative workers for the EU-27. Section 3 demonstrates that there are significant differences in the frequency of creative work activity across nations after adjusting for differences in occupation and industrial structure. Section 4 examines the relationship between creativity and measures of national labour market structures and systems of education and training. Section 5 concludes and briefly alludes to implications for EU policy.

2. MEASURING THE CREATIVE WORKFORCE

A major theme in the behavioural psychology literature on creativity is that of 'eminence' or being 'unique in the whole world', and there are a number of empirical studies of creativity focusing on the lives of truly exceptional musicians, artists or scientists. In contrast to this focus on eminence, there is a body of research focusing on 'everyday' or 'local' creativity of the sort that a large percent of the working population engage in during their daily work activity (Reilly, 2008; Craft, 2005; Richards 1996).

Florida's notion of the creative class corresponds to this latter, more broadly distributed, form of creativity. In Florida (2002), he states that the distinguishing characteristic of the creative class is that its members "engage in work whose function is to create meaningful new forms" (p. 68). The highest level of creative work, characteristic of what Florida refers to as the 'super-creative core', involves "producing new forms or designs that are readily transferable and widely useful..." (p. 69). This group includes such occupations as scientists, university professors, poets and architects. Beyond this core, Florida includes within the creative class a diverse group of professionals who, "engage in creative problem-solving, drawing on complex bodies of knowledge to solve specific problems." He observes, "What they (creative professionals) are required to do regularly is think on their own." (p. 69). Further, he notes that many technicians are included in the creative class as they, "apply complex bodies of knowledge to working with physical materials" and in a number of fields, "are taking on increased responsibility to interpret their work and make decisions..." (pp. 69-70).

As the above references show, Florida's notion of the creative class is an economic one based on the kinds of work activities or jobs that different occupational categories typically undertake. Consistent with this, and in order to measure the size of the creative class and its growth over time for the US economy, he draws on the occupational classifications and figures compiled by the US Bureau of Labor Statistics. Based on standard characterisations of skill requirements for different jobs, occupations are grouped into the creative, service or working classes. The creative class is defined to include most management occupations, professionals and selected categories of technicians and assistant professionals.[3] Florida (2002: 75, 330) estimates that the creative class increased from 3 million workers or 10% of the workforce in 1900 to 38.3 million or 30% in 1999. In 1999, the working and service classes are estimated at 26.1 and 43.4 % of the work force respectively, with agriculture making up the remaining 0.5%.

While this way of measuring the size of the creative workforce is appropriate for capturing broad changes in the importance of creativity within an economy over time, it has a number of limitations that make it unsuitable for an internationally comparative analysis. Firstly, as Florida's discussion of the creative factory emphasises (Florida, 2002: 52), creativity can extend down from the firm's management and technical services to the shop floor, and highly creative firms typically seek to mobilise the knowledge and skills of the entire workforce. The blanket characterisation of the work of operators, sales and service staff and craft workers as non-creative is at odds with a vast literature on 'learning organisations' that emphasize the collective and multi-level nature of learning and creativity at the workplace.

Second, there is a well-established internationally comparative literature which identifies important national effects on the organisation of work. In particular, detailed international comparisons show that the jobs and work activity of the same occupational categories can display significant national variations, requiring more or less learning and problem-solving activity and differences in responsibility and autonomy (Dore, 1973; Gallie, 1978; Lane, 1989; Lorenz and Valeyre, 2005; Maurice *et al.,* 1982). The findings of these detailed international comparisons of enterprise organisation are consistent with the work on national systems of innovation that links the micro-dynamics of learning and knowledge creation to differences in national labour markets and education and training systems.

Third, work on sectoral systems of innovation (Malerba, 2002) points to important differences in the technological dynamism of different sectors of the economy and thus it can be anticipated that the work of the same occupational categories will display marked differences in terms of problem-solving activity and creativeness according to the sector. This implies a need to take into account differences in industrial structure across nations in any statistical analysis of the determinants of creativity at the workplace.

In order to develop a measure of the creative workforce that is suitable for

[3] Florida's measure of the creative class is based on the 1998 Standard Occupational Classification (SOC) system which, in common with the International Standard Classification of Occupations (ISCO), groups jobs together in occupations and more aggregate groups mainly on the basis of the similarity of skills required to fulfil the tasks and duties of the jobs. For the list of occupational categories included in the creative class see Florida (2002: 328-329). In terms of the ISCO used by the European Union, Florida's creative class is composed of Management Occupations (ISCO 12-13), Professionals (ISCO 21-24) and some of the occupations classified as Technicians and Associate Professionals (ISCO 31-34). The service class is composed of Clerks (ISCO 41-42), Service Workers (ISCO 51-52) and Sales and Service Elementary Occupations (ISCO 91). The working class is composed of Craft and Related Trade Workers (ISCO 71-74), Plant and Machine Operators (ISCO 81-83) and Labourers in Mining, Construction, Manufacturing and Transport (ISCO 93).

a comparative analysis of the EU-27, we draw on the results of the 4th EWCS conducted by the European Foundation for the Improvement of Living and Working Conditions in 2005. The survey questionnaire was directed to approximately 1000 active persons in each country with the exception of Cyprus, Estonia, Luxembourg, Malta and Slovenia, which had approximately 600 respondents. The total population surveyed was 29,679 occupied persons.[4] The analysis presented here is restricted to the 9,198 salaried employees working in establishments with at least 10 employees in both industry and services, but excluding agriculture and fishing, public administration and social security, education, health and social work, and private domestic employees. In order to develop a measure of the creative workforce that is consistent with Florida's (2002) characterisation of the creative class, we use factor analysis to identify the underlying associations that exist among six binary variables that capture key features of creative work activity (see Table 1).[5] We then use hierarchical clustering in order to group the population into three basic types of workers: creative workers, constrained problems solvers and workers with Taylorist jobs.[6]

Table 1 shows the percentage of the population characterised by the six work-activity variables that are used in order to classify a worker as creative. Thus, as Florida observes, creative workers typically engage in complex problem-solving activities. Further, workers who use their own ideas in settings where they exercise considerable discretion over their work methods or task order correspond to the distinguishing creative feature of being able to, "think on their own" and to take on, "increased responsibility to interpret their work and make decisions".

These six work activity variables do not, however, provide a basis for discriminating between the creative workforce as a whole and the 'super creative core' which Florida defines in terms of the outcome of producing transferrable and widely-used new forms or designs.

Table 2 shows the composition of the three groups resulting from the hierarchical cluster analysis. The first group, which accounts for 51% of the population, is distinctive for high levels of problem-solving, learning and task complexity. The persons grouped in this cluster use their own ideas and exercise considerable autonomy in carrying out their jobs. We refer to them as 'creative workers'. The second group is characterised by nearly as high levels of problem- solving and learning and comparable levels of task complexity. However, there is little use of one's own ideas, and levels of autonomy or discretion at work are low. This cluster groups employees who, while regularly solving technical or other problems at work, do so in highly supervised settings offering little scope for developing original or creative solutions based on their own ideas. We refer to them as 'constrained problem solvers'. The third group is composed largely of persons doing deskilled work. Levels of learning, problem-solving and task complexity are low. There is little use of one's own ideas and there is limited scope for exercising discretion in how work is carried out. We refer to this group as 'Taylorised workers'.

Table 3 shows that there are variations in the importance of creative learners according to industrial sector. In particular, creative learners are over-represented in business services and in community, social and personal services, while they are under-represented in manufacturing, construction and retail and other services.

[4] The sample of the EWCS is representative of persons in employment (employees and self-employed). The sampling design had the following stages: stratification of primary sampling units (PSUs) according to region and urbanisation level; random selection of starting addresses within each PSU; and a 'random walk' procedure for the selection of the household. The response rate was 0.48 and was calculated as the proportion of completed interviews to the total number of eligible cases. Three types of weighting were applied to the data in order to enhance the representativeness of the results: a selection probability weighting; a non-response (or post-stratification) weighting; and a cross-national weighting in order to be able to do cross-national estimations. For the quality report on the 4th EWCS (see Petrakos Agilis, M., 2007).

[5] For the exact wording of the questions upon which the measures are based, see the Annex; Table A1.

[6] The factor method used is multiple correspondence analysis. In order to group the individuals, Ward's method of hierarchical clustering is used on the basis of the factor scores, or the coordinates of the observations, of the first two factors which account for 59% of the total variance of the data set. See the Annex for a graphical presentation of the factor analysis.

Table 1. Creative work variables

	Percentage of occupied persons affected
Problem solving activities in work	79
Learning new things at work	68
Undertaking complex tasks	62
Using one's own ideas at work	50
Able to choose or change one's work methods	60
Able to choose or change the order of one's tasks	56
N	**9,240**

Source: Fourth Working Conditions Survey, 2005, European Foundation for the Improvement of Living and Working Conditions

Table 2. Cluster analysis of types of workers

	Percentage of occupied persons by type of learner reporting each variable			
Variable	Creative workers	Constrained problem-solvers	Taylorised workers	Average
Problem solving activities at work	96	87	37	79
Learning new things at work	87	84	16	68
Undertaking complex tasks	80	81	8	62
Using one's own ideas at work	77	24	19	50
Able to choose or change one's work methods	94	21	29	60
Able to choose or change the order of one's tasks	92	14	25	56
Total share of occupied persons	**51**	**24**	**25**	**100**

Source: Fourth Working Conditions Survey, 2005, European Foundation for the Improvement of Living and Working Conditions

Table 3. Type of Worker by Sector of Activity

	Percentage of occupied persons by sector of activity and type of learner			
	Creative workers	Constrained problem solvers	Taylorised workers	Total
Manufacturing, construction and utilities	46	27	27	100
Retail and other services	49	23	29	100
Business and financial services	67	19	13	100
Community, social and personal services	59	18	22	100
Average	51	24	25	100

Source: Fourth Working Conditions Survey. European Foundation for the Improvement of Living and Working Conditions

Table 4. Type of Worker by Occupational Category

	Percentage of occupied persons by occupational category and type of worker			
	Creative workers	Constrained problem solvers	Taylorised workers	Total
Senior managers	82	10	7	100
Professionals and technicians	74	18	8	100
Clerks and service workers	53	23	24	100
Skilled workers and machine operators	38	30	32	100
Unskilled workers	33	24	43	100
Average	51	24	25	100

Source: Fourth Working Conditions Survey, European Foundation for the Improvement of Living and Working Conditions

Table 5. National differences in Types of Learners: EU27 (percentage of occupied persons by country and type of learner)

	Creative workers	Constrained problem solvers	Taylorised workers	Total
Belgium	60	21	19	100
Czech Republic	40	30	30	100
Denmark	70	15	14	100
Germany	52	23	26	100
Estonia	58	22	20	100
Greece	39	33	28	100
Spain	35	30	36	100
France	63	18	19	100
Ireland	58	18	24	100
Italy	37	29	34	100
Cyprus	42	26	32	100
Latvia	53	19	27	100
Lithuania	35	27	38	100
Luxembourg	60	20	20	100
Hungary	44	31	25	100
Malta	70	14	16	100
Netherlands	67	16	16	100
Austria	50	28	23	100
Poland	43	34	23	100
Portugal	46	24	29	100
Slovenia	50	25	25	100
Slovakia	33	32	35	100
Finland	66	21	13	100
Sweden	82	10	8	100
United Kingdom	51	22	27	100
Bulgaria	39	30	31	100
Romania	35	38	27	100
EU-27	51	24	25	100

Source: Fourth Working Conditions Survey, 2005 European Foundation for the Improvement of Living and Working Conditions

Table 4 points to the considerable diversity that exists in the importance of creative work across broadly defined occupational categories. Although the large majority of senior managers, professionals and technicians, which make up the bulk of Florida's creative class, are highly over-represented in the creative workers cluster, roughly half of the occupations making up the clerks and sales and service category, who form Florida's service class, engage in creative work activity. Moreover, a significant minority of the manual occupations making up Florida's working class engage in work requiring creative learning, problem-solving and the use of one's own ideas. The results shown in Table 4 point, for the EU at any rate, to the limitations of using standard occupational categories as the basis for identifying the group of workers who are creative at work. Significant proportions of service and manual workers may work in settings where they are called upon to make creative use of their own ideas.

3. NATIONAL EFFECTS

Table 5 identifies important differences in the size of the creative workforce across the EU-27. Creative workers are most present in the Scandinavian countries, Finland, the Netherlands and Malta and least present in Spain, Greece, Italy, Cyprus, the Czech Republic, Hungary, Lithuania, Poland, Slovakia, Bulgaria and Romania. There are intermediate levels of creative work activity in the continental European nations, the UK and Ireland, Portugal and amongst the new member nations in Estonia, Latvia, and Slovenia. The frequency of Taylorised workers tends to show the reverse trend to that of creative workers, being lowest in the Scandinavian countries, the Netherlands and Malta, and highest in the southern nations, Cyprus, the Czech Republic, Hungary, Lithuania, Slovakia, and Bulgaria. The frequency of constrained problem solvers is relatively high in Greece, in the new member nations with the exception of Latvia, and Slovenia and to a lesser extent Lithuania.

Table 3 shows that the frequency of creative work activity varies across broadly defined sectors of activity, and some of the national differences in the importance of creativity shown in Table 5 may be attributable to cross-national differences in industrial structure. It is also possible that international differences in occupational structure may explain some of the differences in the frequency of creative work across the member nations of the EU. In order to control for the effects of these structural variables, we have undertaken a logistic regression analysis explaining the likelihood that an employee is engaged in creative work in terms of nation, industrial sector and occupation. The results are presented in Table 6.

The results presented in the first column of Table 6 show national effects on the likelihood of creative work without structural controls and the second column shows the results with these controls. Germany, which has a profile of types of learners close to the EU-27 average, is used as the reference case. The results thus show whether or not creative work activity is significantly more likely in a nation relative to the German case.

The first column results (without structural controls) show that creative work is significantly more likely in the Nordic countries, Netherlands, Malta and France. Expressed in odds ratios an employee working in Sweden is 4.4 times as likely as an employee in Germany to be engaged in creative work activity. The likelihood of creative work is not significantly different in

Table 6. Logit estimates of national effects on creative work activity

	Logit estimates without structural controls	Logit estimates with structural controls
Belgium	.33	.43*
Czech Republic	-.45**	-.17
Denmark	.81**	.89**
Germany	Reference	
Estonia	.26	.45*
Greece	-.52**	-.47*
Spain	-.70**	-.48*
France	.48**	.51**
Ireland	.24	.06
Italy	-.60**	-.47*
Cyprus	-.40*	-.31
Latvia	.07	.24
Lithuania	-.67**	-.36*
Luxembourg	.34	.21
Hungary	-.29	-.08
Malta	.81**	1.04**
Netherlands	.66**	.60**
Austria	-.08	.06
Poland	-.33*	-.03
Portugal	-.21	.27
Slovenia	-.06	.10
Slovakia	-.77**	-.61**
Finland	.62**	.68**
Sweden	1.47**	1.64**
United Kingdom	-.01	-.20
Bulgaria	-.54**	-.41*
Romania	-.69**	-.32
Manufacturing, construction and utilities	Reference	
Retail and other services		-.16
Business and financial services		.09
Community, social and personal services		.36*
Senior managers		2.29**
Professionals and technicians		1.62**
Clerks and service workers		.68**
Crafts and related trades	Reference	
Operators and assemblers		-.20

* significant at 5% level;
** significant at 1% level.

the continental nations with the exception of France, and the likelihood is significantly lower in the southern nations with the exception of Portugal for which the difference is not significant. Creative work activity is less likely in a number of the new member nations including Lithuania, Poland, Slovakia, the Czech Republic, Bulgaria and Romania. Expressed in odds ratios, the likelihood of creative work is only .46 times as likely in Slovakia as it is in Germany.

The second column results report national effects after taking into account the effects of cross-national differences in the structure of sectors, occupational categories and establishment size. Most of the national effects reported in column one remain the same in terms of the direction of the sign and statistical significance. The exceptions are Belgium and Estonia for which the positive coefficients are now significant at the 5% level, and Cyprus, Poland and Romania for which the negative coefficients are no longer significant at the 5% level or better. Considering the control variables, we can see that relative to manufacturing and mining creative work is more likely in community, social and personal services, while for business and financial services and retail and other services there are no significant differences. As expected, creative work activity is more likely for senior managers and professionals relative to skilled workers and machine operators, and it is also more likely for clerks and sales workers relative to skilled workers and machine operators. Expressed in terms of odds ratios, senior managers are nearly six times as likely to be engaged in creative work activity and professional and technicians are about four times as likely. Sales staff and clerks are about twice as likely as skilled workers and operators to be engaged in creative work activity.

4. CREATIVITY, LABOUR MARKET STRUCTURE AND SYSTEMS OF EDUCATION AND TRAINING

The analysis above has shown that there are significant differences in the importance of creative work activity across the member nations of the European Union. Relatively little attention has been given in the literature to how national-level institutional arrangements may impact on creativity at the workplace. Although creativity at work might be influenced by a wide range of institutional conditions, in this exploratory analysis we focus on a set of complementary institutional arrangements that have received attention in recent work on national innovation systems: the role of broad competence-based systems of education and training; and the role of labour market systems characterised by the combination of high levels of labour market mobility and well developed systems of unemployment protection. (Hall and Soskice, 2001; Lam and Lundvall, 2006; Lorenz and Lundvall, 2006).

4.1. National labour market structure

Hall and Soskice (2001) have argued that fluid labour markets characterised by few restrictions on hiring and firing may impact positively on innovative performance because they allow firms to rapidly bring in new knowledge from the outside and to reconfigure their knowledge bases. New knowledge can support creative outcomes, such as developing new products or services, because it provides a foundation for exploring novel solutions and for learning that extends beyond the firm's existing areas of expertise. In order to explore this hypothesis, we developed a measure of labour market mobility (LABMOB) based on data from the Labour Force Survey on the share of persons by country whose job had started within the last three months. LABMOB is defined as the average of this share over three quarters: the 2nd quarter of 2005 and the 1st and 2nd quarters of 2006.[7]

While fluid labour markets can contribute to a rapid reconfiguration of a firm's knowledge base, as Lam and Lundvall (2006) have observed, labour market mobility is a two-edged sword for the creative firm. Highly creative firms draw their capability from the diverse know-how and practical problem-solving skills embodied in individual experts. Formal professional knowledge may play only a limited role and the expert's problem-solving capabilities have more to do with experience and tacit knowledge generated through interaction, trial-and-error and experimentation. Because these tacit skills cannot be easily codified, the creative firm faces a problem of reproducing what has been learnt into an organizational memory and it is highly vulnerable when it comes to individuals leaving the organisation.

These problems of accumulating and transferring experience-based tacit knowledge take a different form when firms are organised into localised networks and industry clusters. Mobility across organisational borders within industrial clusters contributes to professional and social relationships which provide the 'social capital' and 'information signals' needed to ensure the efficient accumulation and transfer of tacit knowledge in an inter-firm career framework (Saxenian, 1996). We would argue that such professional and social networks are more likely to arise in institutional settings where high levels of labour market mobility are complemented by well-developed systems of unemployment protection and active labour market policies designed to increase the employability of the unemployed. Unemployment protection can encourage individuals to commit themselves to what would otherwise be considered unacceptably risky career paths that are punctuated by transitions between employment and unemployment or part-time employment. Furthermore, unemployment protection combined with active labour market polices can help assure that extended periods of unemployment will not lead individuals to accept downgrading or accept job offers that do not make use of and build on the experience and knowledge they have gained through previous employment.

For these reasons, it can be argued that labour market mobility is more likely to be skill-enhancing in nations with well-developed systems of unemployment protection combined with active labour market policies. In order to capture the development of such systems of flexicurity (FLXSCR), we use a simple indicator constructed by multiplying LABMOB, the measure of labour market mobility, by LMP, total expenditures on active and passive labour market policies as a percentage of GDP.[8]

Figures1 and 2 below present aggregate correlations at the national level showing the relationship between the frequency of creative work activity and the measures of labour market mobility and flexicurity.[9] The results support the view that labour market mobility is more likely to foster skill enhancement and creativity at the workplace when it is combined with well-developed systems of unemployment protection combined with active labour market policies. Figure 1 shows a weak and non-statistically significant positive relationship between the frequency of creative work and the measure of national labour market mobility, while Figure 2 shows a stronger and statistically significant positive relationship between the

[7] The figures are taken from, *Statistics in Focus*, 'Population and Social Conditions', 6/2006, Eurostat.

[8] The labour market expenditure figures are taken from Eurostat's Labour Market Policy data base. Total expenditures are defined as the sum of active and passive expenditure targeted at one of the following: the unemployed, the employed at risk of becoming unemployed and inactive persons who would like to enter the labour market but are disadvantaged in some way. Active measures include expenditure on training, job rotation and job sharing, employment incentives, direct job creation and start-up incentives. Passive measures include expenditure on out of work income maintenance and early retirement.

[9] We are well aware that simple correlations are primitive when it comes to sorting out causalities. In the paper in the Cambridge Journal of Economics (Lorenz and Lundvall, 2010) we pursue a more sophisticated analysis based upon multilevel regression techniques. The results obtained there coincide with those presented below.

frequency of creative work and the measure of flexible security. Denmark, and to a lesser extent Sweden and Finland, stand out for their relatively high levels of flexible security.

4.2 National systems of education and training

Lam and Lundvall (2006) have argued that national systems of education and training can be distinguished by the extent to which they promote an even distribution of competence across occupational categories as opposed to generating wide disparities in skill levels and learning opportunities. Broad competence-based systems are characterised by their concern to balance investment in formal academic education and the production of third-level degrees, with investment in continuous vocational training and with creating possibilities for further training open to all. Such systems are more conducive to decentralised modes of work organisation and favour the forms of interactive learning and the transmission and mobilisation of tacit knowledge that can contribute to creativity at the workplace.

In order to develop measures of the characteristics of national systems of education and training, we used the aggregate data available on Eurostat's electronic database. We used two measures of the level of a nation's investments in formal academic knowledge and skills: the percentage of the population with third-level education (THRDED); and the number of science and engineering graduates as a percentage of the 20-29 aged cohort in 2005 (S&EGRD). In order to capture the breadth of the education and training system and the extent to which value is placed on developing practical job-related skills, we used the results of the 2003 Labour Force Survey module on life-long learning.[10] As an indicator of the overall level of development of further education and training opportunities, we used the figures on the proportion of the labour force involved in any form of education or training during the four weeks prior to the survey. Equality of access to such further education and training is defined as the ratio of the

[10] The figures are available on Eurostat's electronic data base. The Labour Force Survey lifelong learning module distinguishes between formal, non-formal and informal or self-learning. Formal life-long learning is defined as that provided by the degree-conferring institutions of the formal educational system. Non-formal education and training refers to all forms of taught learning, including that provided by employers, that occurs outside the formal degree-conferring educational system. Informal learning refers to self-taught learning including the use of printed materials and on-line computer based learning. For the Eurostat quality report on the lifelong learning module of the LFS, see: http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Adhoc_modules/2003/ExplanatoryNotes/Final_Report_Ahm2003_EN.pdf

Figure 1. % Creative workers by % labour mobility

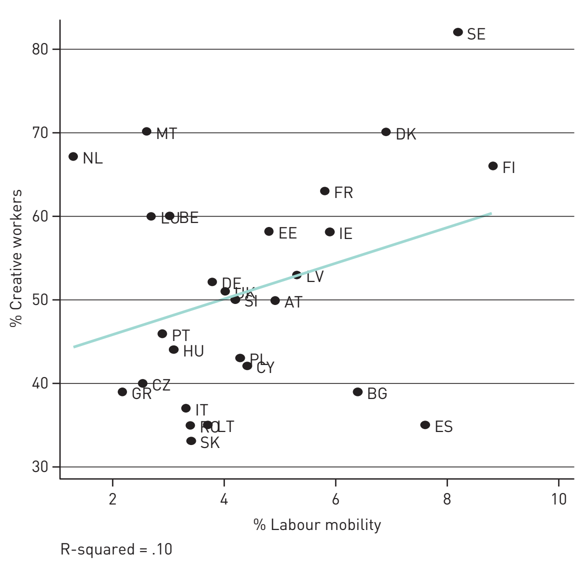
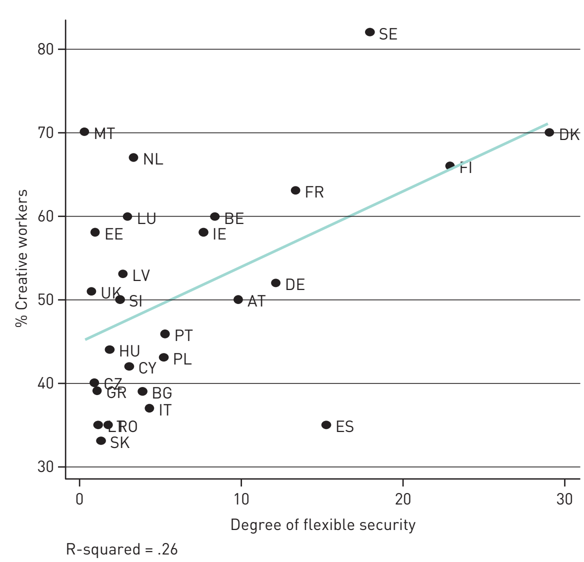


Figure 2. % Creative worker by degree of flexible security



percentage of skilled trades involved in any form of education or training activity to the percentage of managers, professionals and technicians (EQLLL). Higher ratios would indicate a more even distribution of further education and training across occupational categories.[11] The value attached in a nation to developing practical job-related skills and expertise is measured by the proportion of the labour force receiving job-related education or training from providers other than the formal degree-conferring educational system during the four weeks prior to the survey (CVT). This includes continuous vocational training provided by employers.

Figures 3 and 4 show the relationship between the frequency of creative work activity and the two measures of investment in formal academic knowledge. The results are mixed. While there is a positive and statically significant relationship between creativity and the importance of third-level education, there is little discernible relation between creativity at work and the importance of third-level science and engineering degrees.

Figures 5 and 6 show the relationship between creative work activity and our two measures of broad competence-based systems of education and training. The results show a positive and statistically highly significant relationship between the frequency of creative work and the measure of equality of access to life-long learning, and an even stronger and statistically significant positive relationship between the frequency of creative work and a nation's commitment to developing practical job-related skills and expertise.

Overall, while the aggregate correlations suggest that an increase in the resources committed to third-level education in general can promote creativity at work there is little evidence to suggest that insufficient numbers of science and engineering graduates constitute an obstacle to achieving higher levels of creativity at work. Moreover, the results strongly suggest that for many of the new member nations or southern European

Figure 3. % Creative workers by % third level education

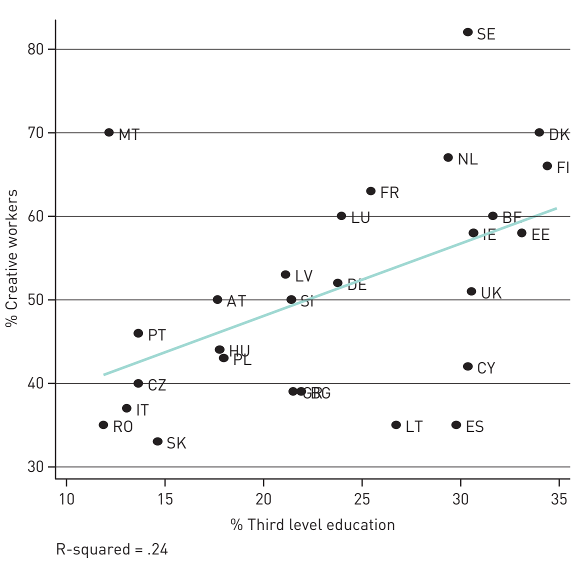
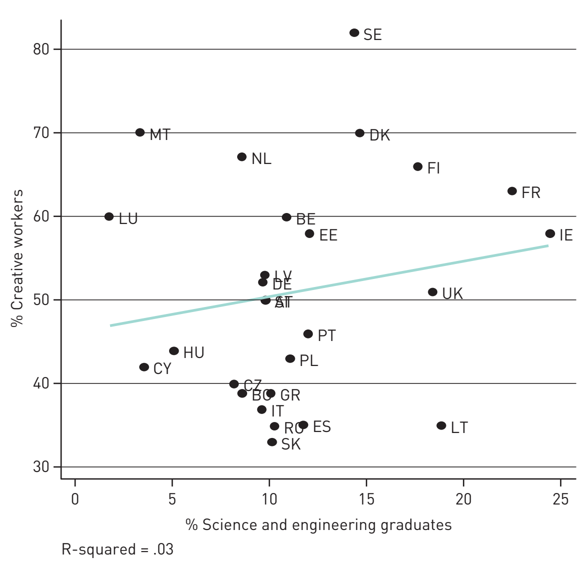


Figure 4. % Creative workers by % science and engineering graduates



[11] We focus on life-long-learning in this analysis rather than on initial vocational training because the further education and training opportunities provided through systems of life-long-learning potentially concern all occupational categories. Apprenticeships or other forms of non-enterprise-based initial vocational training tend to be specific to particular occupations and hence the level of development of these forms of training are less relevant for measuring the overall breadth of the educational and training system in a nation and the extent to which value is placed on the acquisition of job-related skills for employees regardless of their occupational category or level of education.

Figure 5. % Creative workers by degree of equality life long learning

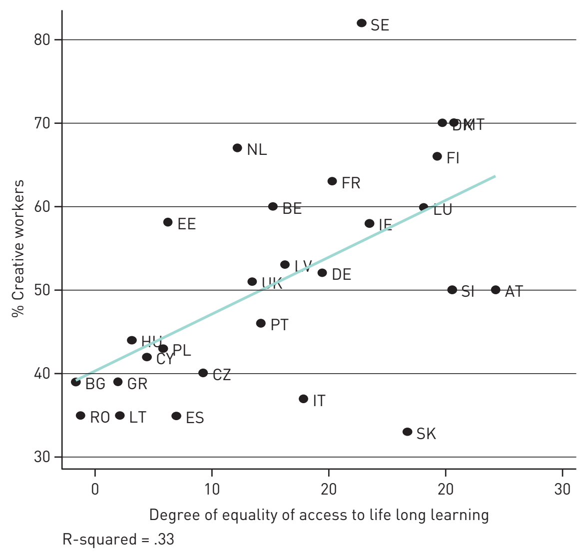
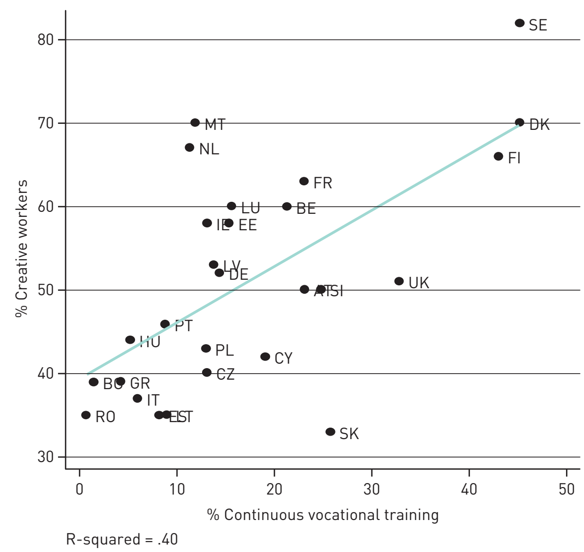


Figure 6. % Creative workers by % continuous vocational training



nations with very low levels of creative work activity the more important bottleneck is their low levels of investment in further education and training.

5. CONCLUSION

One of the starting points for this paper was the analysis of 'the creative class' as developed by Richard Florida. With reference to his work we used employee-level survey data from European countries to classify workers as more or less involved in creative work activity. A first result is that while the occupational categories used by Florida to define the creative class provide a useful means for measuring the rise in the creative class over time within a single nation, they are misleading for purposes of international comparisons. National location matters and we find important 'national effects' on the likelihood of creative work activity after adjusting for the effects of an employee's occupation and sector of activity. A semi or unskilled operator or service worker has a higher chance of being assigned creative tasks in the Scandinavian nations than in the Eastern or Southern nations.

Such national differences may have different causes and in this paper we link them to institutional differences among the European economies that are related to the further education and training system and to labour market structure. We find that broad competence-based education systems characterised by more equal access to further training for enhancing vocational skills as compared to academic knowledge are associated with higher levels of creative work. We also find that labour markets that combine high mobility with ambitious labour market policy in terms of passive and active measures tend to be associated with higher levels of creative work.

Creativity is arguably an essential factor in a nation's capacity for knowledge development and innovation and our analysis has implications for public policy in Europe. Using a somewhat different classification of work categories, Arundel *et al.* (2007) demonstrate that a high frequency of participatory forms of work organisation go hand in hand with more radical innovation within a specific national system. The tendency to operate with narrow perspectives on innovation and innovation systems, where the focus is upon science-driven innovation and neglects experience-based learning and the organisation of work, is misleading (Jensen *et al.*, 2007).

The Spring 2010 European Council emphasised the importance of structural reforms for assuring a strong and sustainable recovery from the current economic crisis and for preserving the sustainability of Europe's social models. Our results provide important guidance for the direction these should take. The analysis shows that 'the creative class' is an elastic category that can be widened to include many ordinary workers. On the basis of our results we would argue that there is a need for institutional reform in Europe in order to create broad-based and egalitarian further education and training systems that are integrated into policies for flexicurity. Such a political mix would serve to spread and deepen creative activities so that they are no longer the privilege of a specific social class. They may also be seen as a way to strengthen the EMU-economies most exposed to global competition and currently under attack from global finance. It would also represent a major step towards aligning the two classical objectives for the Lisbon agenda: competitiveness and social cohesion.

ANNEX

The figure presents graphically the first two axes or factors of the multiple correspondence analysis (MCA). The first factor, accounting for 42% of the variance in the data set or the chi-squared statistic, distinguishes between creative workers and Taylorised workers. On one side of the axis we find the variables measuring the presence of autonomy, learning, problem-solving, complexity and the use of one's own ideas, and on the other side we find the variables measuring their absence. The second factor, accounting for 17% of the variance in the data set, is defined by the presence of problem-solving, learning and complexity combined with the absence of autonomy and the use of one's own ideas at work.

The projection of the centre of gravity of the three worker clusters coming out of the hierarchical classification analysis (see Table 2) onto the graphic representation of the first two factors of the MCA shows that the three clusters correspond to quite different types of work activity. The creative cluster is located to the west of the graph, the Taylorised cluster to the east, and the constrained problem-solving cluster to the south.

Figure A1. Graphical Representation of first two factors of MCA - 6 Organisational Variables. For the six creative work activity variables, + signifies the presence of the feature and – signifies its absence. PBSOLV: problem-solving at work. LEARN: learning new things at work. IDEAS: use of one's own idea at work. COMPLX: undertaking complex tasks. AUTMET: choose or change one's work methods. AUTORD: choose or change one's task order.

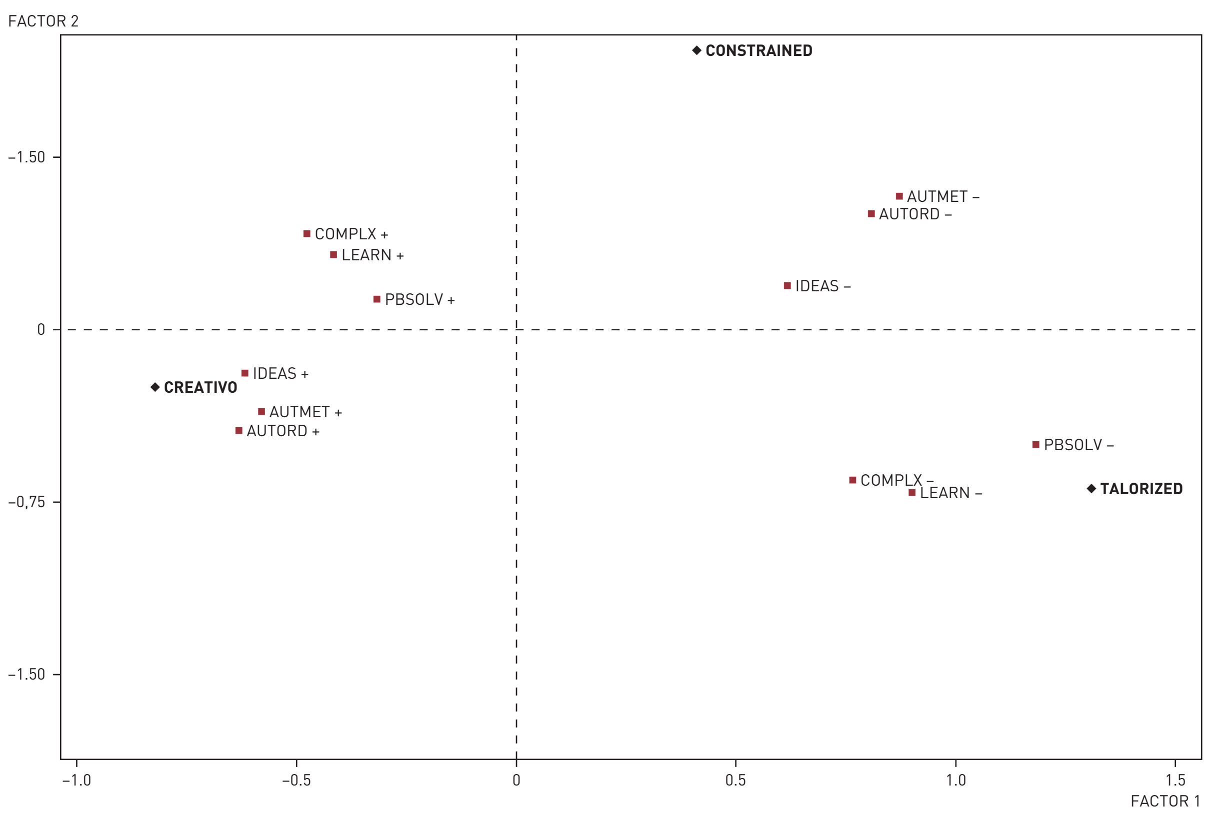


Table A1. Survey questions used for the construction of the binary work activity variables

Variable	Survey questions
Learning new things (LEARN)	Generally, does your main paid job involve, or not, learning new things?
Employee problem-solving (PBSOLV)	Generally, does your main paid job involve, or not, solving unforeseen problems on your own?
Task Complexity (COMPLX)	Generally, does your main paid job involve, or not, complex tasks?
Using one's own ideas in work (IDEAS)	Are you almost always or often able to apply your own ideas in your work?
Discretion in fixing work methods (AUTOMET)	Are you able, or not, to choose or change your methods of work?
Discretion in fixing the order of one's tasks (AUTORD)	Are you able, or not, to choose or change your order of tasks?

Source: Agnès Parent-Thirion, *et al.* 2007, pp. 109-134.

Table A3. Aggregate Indicators: EU-27

	LABMOB	LMP	EQLLL	CVT	THRDED	S&EGRD
Belgium	3.0	2.8	50.5	21.3	31.8	10.9
Czech Republic	2.5	.4	38.5	13.1	13.5	8.2
Denmark	6.9	4.2	79.6	45.2	34.7	14.7
Germany	3.8	3.2	59.0	14.4	23.8	9.7
Estonia	4.8	.2	32.5	15.3	33.3	12.1
Greece	2.2	.5	24.0	4.3	21.5	10.1
Spain	7.6	2.0	34.0	8.2	29.9	11.8
France	5.8	2.3	60.7	23.1	25.5	22.5
Ireland	5.9	1.3	67.1	13.1	30.8	24.5
Italy	3.3	1.3	55.8	6.0	12.9	9.7
Cyprus	4.4	0.7	28.9	19.1	30.5	3.6
Latvia	5.3	0.5	52.6	13.8	21.1	9.8
Lithuania	3.7	0.3	24.3	9.0	26.8	18.9
Luxembourg	2.7	1.1	76.3	15.6	24	1.8
Hungary	3.1	0.6	26.4	5.2	17.7	5.1
Malta	2.6	0.1	81.5	11.9	12	3.4
Netherlands	1.3	2.6	44.5	11.3	29.5	8.6
Austria	4.9	2.0	88.6	23.1	17.6	9.8
Poland	4.3	1.2	31.7	13.0	17.9	11.1
Portugal	2.9	1.8	48.5	8.8	13.5	12
Slovenia	4.2	0.6	81.2	24.8	21.4	9.8
Slovakia	3.4	0.4	73.5	25.7	14.5	10.2
Finland	8.8	2.6	78 .7	43.0	35.1	17.7
Sweden	8.2	2.2	65.6	45.2	30.5	14.4
UK	4	0.2	46.9	32.8	30.7	18.4
Bulgaria	6.4	0.6	16.8	1.5	21.9	8.6
Romania	3.4	0.5	17.6	0.7	11.7	10.3

REFERENCES

AMABILE, T. (1988), "A model of creativity and innovation in organizations", in B. M, Staw and L. L. Cummings (eds.), *Research in Organisational Behaviour*, Greenwich, CT: JAI Press, vol. 10, pp.123-167.

AMABILE, T., R. CONTI, H. COON, J. LAZENBY and M. HERRON (1996), "Assessing the Work Environment for Creativity", *Academy of Management Journal* 39(5), pp. 1154-1184.

ARUNDEL, A., E. LORENZ, B.-Å. LUNDVALL, and A. VALEYRE (2007), "How Europe's Economies Learn: a Comparison of Work Organization and Innovation Mode for the EU-15", *Industrial and Corporate Change*, 16(6), pp. 1175-1210.

BARRON and HARRINGTON (1981), "Creativity, Intelligence, and Personality", *Annual Review of Psychology***,** vol. 32, pp. 439-476.

BHARADWAJ, S., and A. MENON (2000), "Making Innovation Happen in Organizations: Individual Creativity Mechanisms, Organizational Creativity Mechanisms or Both?" *Journal of Product Innovation Management* 17(6), pp. 424 -434.

CRAFT, A. (2005), *Creativity in Schools: Tensions and Dilemmas*, Oxford: Routledge Press.

CSIKSZENTMIHALYI, M. (1988), "Society, culture, and person: A systems view of creativity" in R. Sternberg (ed.), *The Nature of Creativity: Contemporary Psychological Perspectives*, Cambridge: Cambridge University Press.

DORE, R. (1973), *British Factory, Japanese Factory: The Origins of National Diversity in Industrial Relations*, Berkeley: University of California Press.

DRAZIN, G. and KAZANJIAN (1999), "Multilevel Theorizing About Creativity in Organizations: A Sense-making Perspective", *The Academy of Management Review*, 1999, 24(2), pp. 286-307.

FLORIDA, R. (2002), *The Rise of the Creative Class*, New York: Basic Books.

FLORIDA, R., C. MELLANDER and K. STOLARICK (2008), "Inside the black box of regional development—human capital, the creative class and tolerance", *Journal of Economic Geography* 8, pp. 615–649.

FORD, C. (1996), "A Theory of Individual Creative Action in Multiple Social Domains", *The Academy of Management Review* 21(4), pp. 1112-1142.

GALLIE, D., In Search of the New Working Class, Cambridge, Cambridge University Press.

HALL, P., and D. SOSKICE (2001), *Varieties of Capitalism*, Oxford: Oxford University Press.

HENSON, R. (1996),"In Search of the Creative Personality", *Creativity Research Journal* 9, pp. 295-306.

JENSEN, M. B., B. JOHNSON, E. LORENZ, and B.-Å. LUNDVALL (2007), "Forms of Knowledge and Modes of Innovation", *Research Policy* 36(5), June 2007, pp. 680-693.

LAM, A., and B.-Å. LUNDVALL (2006), "The Learning Organisation and National Systems of Competence Building and Innovation" in E. Lorenz and B.-Å. Lundvall (eds.), *How Europe's Economies Learn: Coordinating Competing Models*, Oxford: Oxford University Press.

LANE, C. (1989), *Management and Labour in Europe*, Aldershot: Edward Elgar.

LORENZ, E. and B.-Å. LUNDVALL (eds.) (2006), *How Europe's Economies Learn: Coordinating Competing Models*, Oxford: Oxford University Press.

LORENZ, E., and B.-Å. LUNDVALL (2010), "Accounting for Creativity in the European Union: A Multi-level Analysis of Individual Competence, Labour Market Structure, and Systems of Education and Training" (forthcoming), *Cambridge Journal of Economics*.

LORENZ, E., and A. VALEYRE (2005), "Organisational Innovation, HRM and Labour Market Structure: A Comparison of the EU-15", *Journal of Industrial Relations*, pp. 424-442.

LUNDVALL, B.-Å. (2002), *Innovation, Growth and Social Cohesion: The Danish Model*, Cheltenham: Edward Elgar.

LUNDVALL, B.-Å., and E. LORENZ (2006), "Welfare and Learning in Europe—How to Revitalize the Lisbon Process and Break the Stalemate", in E. Lorenz and B.-Å. Lundvall (eds.), *How Europe's Economies Learn: Coordinating Competing Models*, Oxford: Oxford University Press.

LUBART, T. (1999), "Creativity across cultures", in R. J. Sternberg (ed.), *Handbook of Creativity*, Cambridge: Cambridge University Press.

MALERBA, F. (2002), "Sectoral Systems of Innovation and Production"' *Research Policy* 31, pp. 247–264

MAURICE, M., F. SELLIER, J.-J. SILVESTRE (1986), *The Social Foundations of Industrial Power*, Cambridge, MA: The MIT Press.

MELLANDER, C. and R. FLORIDA (2006), "The Creative Class or Human Capital? - Explaining Regional Development in Sweden", Working Paper Series in Economics and Institutions of Innovation 79, Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies.

PARENT-THIRION, A., E. FERNÁNDEZ MACÍAS, J. HURLEY and G. VERMEYLEN (2007), *Fourth European Working Conditions Survey*, European Foundation for the Improvement of Living and Working Conditions, Dublin.

PAYNE (1990), "The Effectiveness of Research Issues: A review", in M. West and J. Farr (eds.), *Innovation and creativity at work*, Chichester, UK: Wiley. pp. 101-122.

PETRAKOS AGILIS, M. (2007), "Quality Report of the 4th European Working Conditions Survey", European Foundation for the Improvement of Living and Working Conditions, Dublin. http://www.eurofound.europa.eu/publications/htmlfiles/ef0708.htm.

REILLY, R. (2008), "Is Expertise a Necessary Precondition for Creativity? A Case of Four Novice Learning group Facilitators", *Thinking Skills and Creativity* 3, pp. 59-76.

RICHARDS, R. (1996), "Beyond Piaget: Accepting Divergent, Chaotic, and Creative Thought", *New Direction for Child Development*, 72, pp. 67-78.

ROFFE, I. (1999), "Innovation and Creativity in Organisations: a Review of the Implications for Training and Development", *Journal of European Industrial Training* 23, pp. 224-241.

SAXENIAN, A. (1996), *Beyond Boundaries: Open Labor Markets and Learning in Silicon Valley,* Oxford: Oxford University Press.

STERNBERG, R. (ed.) (1988), *The Nature of Creativity: Contemporary Psychological Perspectives*, Cambridge: Cambridge University Press.

STERNBERG, R. J. (ed.) (1999), *Handbook of Creativity*, Cambridge: Cambridge University Press.

STERNBERG, R., and T. Lubart (1991), "An Investment Theory of Creativity and its Development"', *Human Development* 34(1), pp. 1-31.

WEISBERG, R. (1993), *Creativity: Beyond the Myth of Genius*, New York: W.H. Freeman.

WHITLEY, R. (2006), "Innovation Systems and Institutional Regimes: the Construction of Different Types of National, Sectoral and Transnational Innovation systems"' in E. Lorenz and B.-Å. Lundvall (eds.), *How Europe's Economies Learn: Coordinating Competing Models*, Oxford: Oxford University Press.

WOODMAN, R. W., J. E. SAWYER and R. W. GRIFFIN (1993), "Toward a Theory of Organizational Creativity", *Academy of Management Review* 18(2), pp. 293-321.

BBVA

Innovative Organizations: Structure, Learning and Adaptation

Alice Lam
School of Management Royal Holloway University of London

1. INTRODUCTION

Innovation is an important source of growth and a key determinant of competitive advantage for many organizations. Achieving innovation requires the coordinated efforts of many different actors and the integration of activities across specialist functions, knowledge domains and contexts of application. Thus, organizational creation is fundamental to the process of innovation (Van de Ven et al 1999). The ability of an organization to innovate is a pre-condition for the successful utilization of inventive resources and new technologies. Conversely, the introduction of new technology often presents complex opportunities and challenges for organizations, leading to changes in managerial practices and the emergence of new organizational forms. Organizational and technological innovations are intertwined. Schumpeter (1950) saw organizational changes, alongside new products and processes, as well as new markets as factors of 'creative destruction'.

Extant literature on organizational innovation is very diverse and can be broadly classified into three streams. Organizational design theories focus predominantly on the link between structural forms and the propensity of an organization to innovate (e.g. Burns and Stalker, 1961; Lawrence and Lorsch, 1967; Mintzberg, 1979). The unit of analysis is the organization and the main research aim is to identify the structural characteristics of an innovative organization, or to determine the effects of organizational structural variables on product and process innovation. Theories of organizational cognition and learning (Glynn, 1996; Bartel and Garud, 2009), by contrast, emphasise the cognitive foundations of organizational innovation which is seen to relate to the learning and organizational knowledge creation process (Agyris and Schon, 1978; Nonaka and Takeuchi, 1995; Nonaka and von Krogh, 2009). This strand of work provides a micro-lens for understanding the capacity of organizations to create and exploit new knowledge necessary for innovative activities. A third strand of research concerns organizational change and adaptation, and the processes underlying the creation of new organizational forms (Lewin and Volberda, 1999). Its main focus is to understand whether organizations can adapt in the face of radical environmental shifts and technological change. In this context, innovation is considered as a capacity to respond to changes in the external environment, and to influence and shape it (Burgleman, 1991; 2002; Child, 1997; Teece, 2007).

This chapter examines the nature of innovative organizations and the relationship between organizing and innovating from these three perspectives. Section two will draw on organizational design theories and work in the field of strategy to examine the relationship between organizational structure and innovativeness. The third section looks at the micro-level process of organizational learning and knowledge creation. It argues that organizations with different structural forms vary in their patterns of learning and knowledge creation, engendering different types of innovative capabilities. This will be followed by an analysis of organizational adaptation and the contemporary challenges facing firms in pursuing 'organizational ambidexterity' for sustaining innovation. The final section draws some general conclusions from the analysis and highlights the gaps in the existing literature and areas for future research.

2. ORGANIZATIONAL STRUCTURE AND INNOVATION

2.1. Structural archetypes and innovativeness

The classical theory of organizational design was marked by a preoccupation with universal forms and the idea of 'one best way to organise'. The work of Weber (1947) on the bureaucracy and of Chandler (1962) on the multidivisional form, was most influential. The assumption of 'one best way' was, however, challenged by research carried out during the 1960s and 1970s under the rubric of contingency theory which explains the diversity of organizational forms and their variations with reference to the demands of context. Contingency theory argues that the most 'appropriate structure' for an organization is the one that best fits a given operating contingency, such as scale of operation (Blau, 1970), technology (Woodward, 1965; Perrow, 1970) or environment (Burns and Stalker, 1961; Lawrence and Lorsch, 1967). This strand of research and theory underpins our understanding of the relationships between the nature of the task and technological environments, structure and performance. Some of the studies deal specifically with the question of how structure is related to innovation.

Burns and Stalker's (1961) polar typologies of 'mechanistic' and 'organic' organizations (see Box 1) demonstrate how the differences in technological and market environment, in terms of their rate of change and complexity, affect organizational structures and innovation management. Their study found that firms could be grouped into one of the two main types: the former more rigid and hierarchical, suited to stable conditions; and the latter, a more fluid set of arrangements, adapting to conditions of rapid change and innovation. Neither type is inherently right or wrong, but the firm's environment is the contingency that prompts a structural response. Related is the work of Lawrence and Lorsch (1967) on principles of organizational differentiation and integration and how they adapt to different environmental conditions, including the market -- technical-economic and the scientific sub-environments, of different industries. Whereas Burns and Stalker treat an organization as an undifferentiated whole that is either mechanistic or organic, Lawrence and Lorsch recognize that mechanistic and organic structures can co-exist in different parts of the same organization owing to the different demands of the functional sub-environments. The work of these earlier authors had a profound impact on organizational theory and provided useful design guidelines for innovation management. Burns and Stalker's model remains highly relevant for our understanding of the contemporary challenges facing many organizations in their attempts to move away from the mechanistic towards the organic form of organizing, as innovation becomes more important and the pace of environmental change accelerates. Lawrence and Lorsch's suggestion that mechanistic and organic structures can coexist is reflected in the contemporary debate about the importance of developing hybrid modes of organizations—'ambidextrous organizations'—that are capable of coping with both evolutionary and revolutionary technological changes (O'Reilly and Tushman, 2004; 2008; Tushman *et al.,* 2010; see section 4).

Another important early contribution is the work of Mintzberg (1979) who synthesised much of the work on organizational structure and proposed a series of archetypes that provide the basic structural configurations of firms operating in different environments. In line with contingency theory, he argues that the successful organization designs its structure to match its situation. Moreover, it develops a logical configuration of the design parameters. In other words, effective structuring requires consistency of design parameters and contingency factors. The

Table 1. Burns and Stalker: Mechanistic and Organic Structures

Burns and Stalker set out to explore whether differences in the technological and market environments affect the structure and management processes in firms. They investigated twenty manufacturing firms in depth, and classified environments into 'stable and predictable' and 'unstable and unpredictable'. They found that firms could be grouped into one of the two main types, mechanistic and organic forms, with management practices and structures that Burns and Stalker considered to be logical responses to environmental conditions.

The *Mechanistic Organization* has a more rigid structure and is typically found where the environment is stable and predictable. Its characteristics are:

a. tasks required by the organization are broken down into specialised, functionally differentiated duties and individual tasks are pursued in an abstract way, that is more or less distinct from the organization as a whole;

b. the precise definition of rights, obligations and technical methods is attached to roles, and these are translated into the responsibilities of a functional position; there is also a hierarchical structure of control, authority and communication;

c. knowledge of the whole organization is located exclusively at the top of the hierarchy, with greater importance and prestige being attached to internal and local knowledge, experience and skill rather than that which is general to the whole organization;

d. there is a tendency for interactions between members of the organization to be vertical, i.e. between superior and subordinate.

The *Organic Organization* has a much more fluid set of arrangements and is an appropriate form for changing environmental conditions which require emergent and innovative responses. Its characteristics are:

a. individuals contribute to the common task of the organization and there is continual adjustment and re-definition of individual tasks through interaction with others;

b. there is spread of commitment to the organization beyond any technical definition, a network structure of control authority and communication, and the direction of communication is lateral rather than vertical;

c. knowledge may be located anywhere in the network, with this ad hoc location becoming the centre of authority and communication;

d. importance and prestige attach to affiliations and expertise valid in industrial, technical and commercial milieus external to the firm.

Mechanistic and organic forms are polar types at the opposite ends of a continuum and, in some organizations, a mixture of both types can be found.

Source: Burns and Stalker (1961).

'configurational hypothesis' suggests that firms are likely to be dominated by one of the five pure archetypes identified by Mintzberg, each with different innovative potential: simple structure, machine bureaucracy, professional bureaucracy, divisionalised form and adhocracy. Two of these archetypes can be classified as organic organizations with a high capacity for innovation and adaptation: the simple structure and the adhocracy. The former relies on direct supervision by one person, as in the case of entrepreneurial start-ups, which continuously searches high-risk environments. The latter is a highly flexible project-based organization relying on the mutual adjustment of problem-solving teams. It is capable of radical innovation in a volatile environment. The other three remaining archetypes, machine bureaucracy, professional bureaucracy and the divisionalized form are more inhibited in their innovative capabilities and less able to cope with novelty and change. The characteristic features of the archetypes and their innovative implications are shown in Table 1.

Contingency theories account for the diversity of organizational forms in different technological and task environments. They assume that as technology and product markets become more complex and uncertain, and task activities more heterogeneous and unpredictable, organizations will adopt more adaptive and flexible structures, and they will do so by moving away from bureaucratic to organic forms of organizing. The underlying difficulties in achieving the 'match', however, are not addressed in this strand of research.

2.2. Strategy, structure and the innovative firm

The work of micro-economists in the field of strategy considers organizational structure as both cause and effect of managerial strategic choice in response to market opportunities. Organizational forms are constructed from the two variables of 'strategy' and 'structure'. The central argument is that certain organizational types or attributes are more likely to yield superior innovative performance in a given environment because they are more suited to reduce transaction costs and cope with potential capital market failures. The multi-divisional, or M-form, for example, has emerged in response to increasing scale and complexity of enterprises and is associated with a strategy of diversification into related product and technological areas (Chandler, 1962). It can be an efficient innovator within certain specific product markets, but may be limited in its ability to develop new competencies.

Lazonick's theory of 'the innovative enterprise' (Lazonick, 2005; 2010) is rooted in the Chandlerian framework, inasmuch as it focuses on how strategy and structure determine the competitive advantage of the business enterprise. It also builds on Lawrence and Lorsch's (1967) conceptualisation of organizational design problems as differentiation and integration. The theory distinguishes the 'optimizing firm' from the 'innovative firm'. While the former seeks to maximize profits within given technological capabilities and market constraints, the latter seeks to transform technological and markets constraints through the development of distinctive organizational capabilities which cannot be easily imitated by competitors. Lazonick identifies three social conditions that support the development of the innovative firm. The first condition is 'strategic control' which refers to the set of relations that gives key decision-makers the power, knowledge and incentives to allocate the firm's resources to confront market threats and opportunities.

Table 2. Mintzberg's structural archetypes and their innovative potentials

Organization archetype	Key features	Innovative potential
Simple structure	An organic type centrally controlled by one person, which can respond quickly to changes in the environment, e.g. small start-ups in high-technology.	Entrepreneurial and often highly innovative, continually searching for high-risk environments. Weaknesses are the vulnerability to individual misjudgement and resource limits on growth.
Machine bureaucracy	A mechanistic organization characterized by a high level of specialization, standardization and centralized control. A continuous effort to routinize tasks through formalization of worker skills and experiences, e.g. mass production firms.	Designed for efficiency and stability. Good at dealing with routine problems, but highly rigid and unable to cope with novelty and change.
Professional bureaucracy	A decentralised mechanistic form which accords a high degree of autonomy to individual professionals. Characterized by individual and functional specialization, with a concentration of power and status in the 'authorized experts'. Universities, hospitals, law and accounting firms are typical examples.	The individual experts may be highly innovative within a specialist domain, but the difficulties of coordination across functions and disciplines impose severe limits on the innovative capability of the organization as a whole.
Divisionalized form	A decentralized organic form in which quasi-autonomous entities are loosely coupled together by a central administrative structure. Typically associated with larger organizations designed to meet local environmental challenges.	An ability to concentrate on developing competency in specific niches. Weaknesses include the 'centrifugal pull' away from central R&D towards local efforts, and competition between divisions which inhibit knowledge sharing.
Adhocracy	A highly flexible project-based organization designed to deal with instability and complexity. Problem-solving teams can be rapidly reconfigured in response to external changes and market demands. Typical examples are professional partnerships and software engineering firms.	Capable of fast learning and unlearning; highly adaptive and innovative. However, the unstable structure is prone to short life, and may be driven over time toward bureaucracy (see also section 3.2).

Sources: Mintzberg (1979); Tidd *et al.* (1997: 313-314); Lam (2000).

The second condition is 'organizational integration' -- that is the horizontal and vertical integration of skills and knowledge to support cumulative learning over time. And the third condition is 'financial commitment' to ensure that sufficient funds are allocated for competence development to sustain the cumulative innovation process. The essence of the innovative enterprise, according to Lazonick (2005: 34), "is the organizational integration of a skill base that can engage in collective and cumulative learning". The critical importance of skills and knowledge integration as the social foundations of innovation is also stressed by several other authors (Lam, 2000; Lam and Lundvall, 2006).

Because the conditions that underpin the innovative firm are social, the type of organisational integrative capability and the nature of the innovative firm tend to vary across institutional contexts and over time (Whitley, 2000; Lazonick, 2005). Drawing on comparative historical evidence, Lazonick (2005) has illustrated the rise and fall of different national models of innovative firms characterised by different types of organizational capabilities. For example, the growth of the US industrial corporation during the first half of the twentieth century was energised by a powerful managerial organization for deploying new technology and using unskilled and semi-skilled workers in mass production. The US managerial corporation was confronted by the Japanese model of the innovative firm in the 1970s which outperformed the US in many industrial sectors such as consumer electronics, machine tools and automobiles. Japanese firms have been able to gain a competitive advantage in these industries because of their superior organizational capacity for integrating shop-floor skills and enterprise networks, enabling them to plan and coordinate specialised divisions of labour and innovative investment strategies. The late 1990s saw the resurgence of the US high-technology sectors spearheaded by what Lazonick (2005; 2010) refers to as 'new economy companies' in Silicon Valley which drew their innovative capabilities from the dynamic integration of technical and entrepreneurial skills within highly flexible, open network organizations.

The theory of the innovative firm propounded by Lazonick, alongside other researchers in the field of strategy (e.g. Teece *et al.,* 1997; Teece, 2007) stresses the importance of organizational and managerial processes—integrating, learning and reconfiguration—as core elements that underpin firms' innovative performance. However, this strand of work devotes little attention to the micro-dynamics of learning within organizations.

3. THE COGNITIVE FOUNDATIONS OF ORGANIZATIONAL INNOVATION

3.1. Organizational learning and knowledge creation

The structural perspectives discussed above treat innovation as an output of certain structural features. Some organizational researchers regard innovation as a process of bringing new, problem-solving, ideas into use (Amabile, 1988; Kanter, 1983). Mexias and Glynn (1993: 78) define innovation as "non-routine, significant, and discontinuous organizational change that embodies a new idea that is not consistent with the current concept of the organization's business". This approach defines an innovative organization as one that is intelligent and creative (Glynn, 1996; Woodman *et al.,* 1993), capable of learning effectively (Argote, 1999; Senge, 1990; Agyris and Schon, 1978) and creating new knowledge (Nonaka, 1994; Nonaka and Takeuchi, 1995; Nonaka and von Krogh, 2009). Cohen and Levinthal (1990) argue that innovative outputs depend on the prior accumulation of knowledge that enables innovators to assimilate and exploit new knowledge. From this perspective, understanding the role of organizational learning in fostering or inhibiting innovation becomes crucially important.

Central to theories of organizational learning and knowledge creation is the question of how organizations translate individual insights and knowledge into collective knowledge and organizational capability. While some researchers argue that learning is essentially an individual activity (Simon, 1991; Grant, 1996), most theories of organizational learning stress the importance of collective knowledge as a source of organizational capability. Collective knowledge is the accumulated knowledge of the organization stored in its rules, procedures, routines and shared norms which guide the problem-solving activities and patterns of interaction among its members. Collective knowledge resembles the 'memory' or 'collective mind' of the organization (Walsh and Ungson, 1991). It can either be a 'stock' of knowledge stored as hard data, or represent knowledge in a state of 'flow' emerging from interaction. Collective knowledge exists between rather than within individuals. It can be more, or less, than the sum of the individuals' knowledge, depending on the mechanisms that translate individual into collective knowledge (Glynn, 1996). Both individuals and organizations are learning entities. All learning activities, however, take place in a social context, and it is the nature and boundaries of the context that make a difference to learning outcomes.

Much of the literature on organizational learning points to the importance of social interaction, context and shared cognitive schemes for learning and knowledge creation (Agyris and Schon, 1978; Lave and Wenger, 1991; Brown and Duguid, 1991, 1998; Bartel and Garud, 2009). This builds on Polanyi's (1966) idea that a large part of human knowledge is subjective and tacit, and cannot be easily codified and transmitted independently of the knowing subject. Hence its transfer requires social interaction and the development of shared understanding and common interpretive schemes.

Nonaka's theory of organizational knowledge creation is rooted in the idea that shared cognition and collective learning constitute the foundation of organizational knowledge creation (Nonaka, 1994; Nonaka and Takeuchi, 1995; Nonaka and von Krogh, 2009). At the heart of the theory is the idea that tacit knowing constitutes the origin of all human knowledge, and organizational knowledge creation is a process of mobilising individual tacit knowledge and fostering its interaction with the explicit knowledge base of the firm. Nonaka argues that knowledge needs a context to be created. He uses the Japanese word 'ba', which literally means 'place', to describe such a context. *'Ba'* provides a shared social and mental space for the interpretation of information, interaction and emerging relationships that serves as a foundation for knowledge creation. Participating in a *'ba'* means transcending one's limited cognitive perspective or social boundary to engage in a dynamic process of knowledge sharing and creation. In a similar vein, the notion of 'community of practice' (Lave and Wenger, 1991; Wenger, 1998; Brown and Duguid, 1991; 1998) suggests that organizational members construct their shared identities and perspectives through 'practice', that is shared work experiences. Practice provides a social activity in which shared perspectives and cognitive repertoires develop to facilitate knowledge sharing and transfer. Hence, the work group provides an important site where intense learning and knowledge creation may develop. The group, placed at the intersection of horizontal and vertical flows of knowledge within the organization, serves as a bridge between the individual and organization in the knowledge creation process. Much of the recent literature on new and innovative forms of organization also focuses on the use of decentralised, group-based structure as a key organizing principle.

Many organizational and management researchers regard the firm as a critical social context where collective learning and knowledge creation take place. Nonaka and Takeuchi (1995) talk about the 'knowledge-creating company'. Argyris and Schon (1978) suggest that an organization is, at its root, a cognitive enterprise that learns and develops knowledge. 'Organizational knowledge' essentially refers to the shared cognitive schemes and distributed common understanding within the firm that facilitate knowledge sharing and transfer. It is similar to Nelson and Winter's (1982) concept of 'organizational routines': a kind of collective knowledge rooted in shared norms and beliefs that aids joint-problem solving and is capable of supporting complex patterns of action in the absence of written rules. The notion of 'core competence' (Prahalad and Hamel, 1990) implies that the learning and knowledge creation activities of firms tend to be cumulative and path-dependent. Firms tend to persist in what they do because learning and knowledge are embedded in social relationships, shared cognition and existing ways of doing things (Kogut and Zander, 1992). Several authors have analysed how collective learning in technology depends on firms' cumulative competences and evolves along specific trajectories (Dosi, 1988; Pavitt, 1991). Thus, the shared context and social identity associated with strong group-level learning and knowledge accumulation processes may constrain the evolution of collective knowledge. Firms may find it difficult to unlearn past practices and explore alternative ways of doing things. Levinthal and March (1993) argue that organizations often suffer from 'learning myopia', and have a tendency to sustain their current focus and accentuate their distinctive competence: what they call falling into a 'competency trap'. The empirical research by Leonardo-Barton (1992) illustrates how firms' 'core capabilities' can turn into 'core rigidities' in new product development.

An inherent difficulty in organizational learning is the need to maintain an external boundary and identity while at the same time keeping the boundary sufficiently open to allow the flow of new knowledge and ideas from outside. March (1991) points out that a fundamental tension in organizational learning is balancing the competing goals of 'the exploitation of old certainties' and the 'exploration of new possibilities'. Whereas knowledge creation is often a product of an organization's capability to recombine existing knowledge and generate new applications from its existing knowledge base, radically new learning tends to arise from contacts with those outside the organization who are in a better position to challenge existing perspectives and paradigms. Empirical research has suggested that sources of innovation often lie outside an organization (von Hippel, 1988; Lundvall, 1992). External business alliances and network relationships, as well as using new personnel to graft new knowledge onto the existing learning systems, are important mechanisms for organizational learning and knowledge renewal in an environment characterised by rapid technological development and disruptive changes (Powell, 1998; Lam, 2007). The 'dynamic capability' perspective argues that the long-term competitive performance of the firm lies in its ability to build and develop firm-specific capability and, simultaneously, to renew and re-configure its competences in response to an environment marked by 'creative destruction' (Teece *et al.*, 1997; Teece 2007). Thus, a fundamental organizational challenge in innovation is not simply the maintenance of a static balance between exploitation and exploration, or stability and change, but a continuous need to balance and coordinate the two dynamically throughout the organization.

3.2. Two alternative models of learning organizations: 'J-form' vs. 'Adhocracy'

All organizations can learn and create knowledge, but their learning patterns and innovative capabilities vary (Lam, 2000; 2002). During the past two decades, an extensive literature has examined new organizational models and concepts designed to support organizational learning and innovation. These models include 'high performance work systems' or 'lean production' (Womack *et al.,* 1990), pioneered by Japanese firms in the automobile industry; and the 'N-form corporation' (Hedlund, 1994) and 'hypertext organization' (Nonaka and Takeuchi 1995). More recently, concepts such as 'cellular forms' (Miles *et al.,* 1997); 'modular forms' (Galunic and Eisenhardt, 2001); 'project-based networks' (DeFillippi, 2002) and 'new economy firms' (Lazonick, 2005) reflect the growth of flexible and adaptive forms of organization with a strategic focus on entrepreneurship and radical innovation in knowledge-intensive sectors of the economy. These studies highlight the different ways in which firms seek to create learning organizations capable of continuous problem solving and innovation.

A closer examination of the literature on new forms suggests that the various models of learning organizations can be broadly classified into two polar ideal types, namely, the 'J-form' and 'adhocracy' (Lam, 2000; 2002). The former refers to an organization which is good at exploitative learning and derives its innovative capabilities from the development of organization-specific collective competences and problem-solving routines. The term J-form is used because its archetypal features are best illustrated by the 'Japanese type' of organizations, such as Aoki's (1988) model of the 'J-firm', and Nonaka and Takeuchi's (1995) 'knowledge creating companies'. Adhocracy (Mintzberg, 1979), by contrast, tends to rely more upon individual specialist expertise organized in flexible market-based project teams capable of speedy responses to changes in knowledge and skills, and integrating new kinds of expertise to generate radical new products and processes. It is skilled at explorative learning. Mintzberg's term is used here to capture the dynamic, entrepreneurial and adaptive character of the kind of organization typified by Silicon- Valley-type companies (Bahrami and Evans, 2000). Both the 'J-form' and 'adhocracy' are learning organizations with strong innovative capabilities, but they differ markedly in their knowledge configurations, patterns of learning and the type of innovative competences generated. These two polar organizational types are facilitated by different institutional characteristics of labour markets and systems of competence building (Lam, 2000; Lam and Lundvall, 2006).

The J-form organization relies on knowledge that is embedded in its operating routines, team relationships and shared culture. It is facilitated by a relatively stable, long-term employment relationship and, a broad-based education and training system for the majority of the workforce. Learning- and knowledge-creation within the J-form takes place within an 'organizational community' that incorporates shopfloor skills in problem solving, and intensive interaction and knowledge sharing across different functional units. The existence of stable organizational careers rooted in an internal labour market provide an incentive for organizational members to commit to organizational goals and to develop firm-specific problem-solving knowledge for continuous product and process improvement. New knowledge is generated through the fusion, synthesis and combination of the existing knowledge base. The J-form tends to develop a strong orientation towards pursuing an incremental innovation strategy and do well in relatively mature technological fields characterised by rich possibilities of combinations and incremental improvements of existing components and products (e.g. machine-based industries, electronics components and automobiles). But the J-form's focus on nurturing organizationally-embedded, tacit knowledge and its emphasis on continuous improvement in such knowledge can inhibit learning radically new knowledge from external sources. The disappointing performance of Japanese firms in such fields as software and biotechnology in the 1990s may constitute evidence of the difficulties faced by 'J-form firms' in entering and innovating in rapidly developing new technological fields (Whitely, 2003).

An adhocracy is an organic and adaptive form of organization that is able to fuse professional experts with varied skills and knowledge into adhoc project teams for solving complex and often highly uncertain problems. Learning and knowledge creation in an adhocracy occurs within professional teams that often are composed of employees from different organizations. Careers are usually structured around a series of discrete projects rather than advancing within an intra-firm hierarchy. The resulting project-based career system is rooted in a relatively fluid occupational labour market which permits the rapid reconfiguration of human resources to align with shifting market requirements and technological changes. The adhocracy has a much more permeable organizational boundary that allows the insertion of new ideas and knowledge from outside. This occurs through the recruitment of new staff, and the open professional networks of the organizational members that span organizational boundaries. The adhocracy derives its competitive strength from its ability to reconfigure the knowledge base rapidly to deal with high levels of technical uncertainty, and to create new knowledge to produce novel innovations in emerging new industries. It is a very adaptive form of organization capable of dynamic learning and radical innovation. However, the fluid structure and speed of change may create problems in knowledge accumulation, since the organization's competence is embodied in its members' professional expertise and market-based know-how which are potentially transferable. The adhocracy is subject to knowledge loss when individuals leave the organization. The long-term survival of this loose, permeable organizational form requires the support of a stable social infrastructure rooted in a wider occupational community or localised firm networks

Although firms in the high-technology sectors are under intense pressure to learn faster and organize more flexibly, evidence thus far suggests that complete adhocracies remain rare. Adhocracies are usually confined to organizational subunits engaged in creative work (e.g. 'skunk work' adhocracies) (Quinn, 1992), or knowledge-intensive professional service fields (e.g. law, management consultancies, software engineering design) where the size of the firm is generally relatively small, enabling the whole organization to function as an interdependent network of project teams (DeFillippi, 2002). Attempts by large corporations to adopt the adhocracy mode have proved difficult to sustain in the long-run (Foss, 2003). Elsewhere, the most successful examples of adhocracies are found in regionally based industrial communities, as in the case of Silicon Valley, and other high-technology clusters (Saxenian, 1996; Angels, 2000). There, the agglomeration of firms creates a stable social context and shared cognitive framework to sustain collective learning and reduce uncertainty associated with swift formation of project teams and organizational change.

4. ORGANIZATIONAL CHANGE AND ADAPTATION: TOWARDS 'ORGANIZATIONAL AMBIDEXTERITY'

Can organizations change and survive in the face of major environmental shifts? If so, how do they adapt? There are two broad perspectives in the research on organizational change. Organizational ecology and institutional theorists (Hannan and Freeman, 1984; Barnett and Caroll, 1995; DiMaggio and Powell, 1983; Greenwood and Hinings, 1996) emphasize the powerful forces of organizational inertia and argue that individual organizations seldom succeed in making radical change in strategy and structure in the face of environmental turbulence. This strand of research focuses on the way environments select organizations, and how this selection process creates change in organizational forms as new entrants within an industry display the established organizations that cannot adapt fast enough. One possible way for organizations to adapt, according to the selectionist perspective, is to spin out new business ventures (Barnett and Freeman, 2001; Christensen, 1997). By contrast, theories of strategic organizational adaptation and change focus on the role of managerial action and strategic choice in shaping organizational change (Child, 1997; Burgleman, 2002; Teece, 2007). They view organizational change as a product of an actor's decisions and learning, rather than the outcome of a passive environmental selection process. According to Child (1997), organizational action is bounded by the cognitive, material and relational structures internal and external to the organization, but at the same time it impacts upon those structures. Organizational actors, through their actions and 'enactment' (Weick, 1979), are capable of redefining and modifying structures in ways that will open up new possibilities for future action. As such, the strategic choice perspective projects the possibility of creativity and innovative change within the organization.

Many strategic adaptation theorists view organizational change as a continuous process encompassing the paradoxical forces of continuity and change. Continuity maintains a sense of identity for organizational learning (Weick, 1996), provides political legitimacy, and increases the acceptability of change among those who have to live with it (Child and Smith, 1987). Burgleman's (1991, 2002) study of Intel corporation illustrates how the company successfully evolved from a memory to a microprocessor company by combining the twin elements of continuity and change for strategic renewal. Burgleman argues that consistently successful organizations use a combination of 'induced' and 'autonomous' processes in strategy-making to bring about organizational renewal. The induced process develops initiatives that are within the scope of the organization's current strategy and build on existing organizational learning (i.e. continuity). In contrast, the autonomous process concerns initiatives that emerge outside the organization and provide the opportunities for new organizational learning (i.e. change). These twin processes are considered vital for successful organizational transformation. In a similar vein, Brown and Eisenhardt (1997) note that continuous organizational change for rapid product innovation is becoming a crucial capability for firms operating in high-velocity industries with short product cycles. Based on case studies of multi-product innovations in the computer industry, the authors conclude that continuous change and product innovations are supported by organizational structures that can be described as 'semi-structures', a combination of 'mechanistic' and 'organic' features, that balance order and chaos.

The dual search for stability and change constitutes a central paradox in all forms of organizing and poses a major challenge for firms operating in today's business environment (Farjoun, 2010). In the past, many organizational theorists maintained that the structures, processes and practices that support stability and reliability were largely incompatible with those needed for change and flexibility. The tension between 'exploitation' and 'exploration' in organizational learning and innovation is a familiar example (March, 1991). Exploitation builds on existing knowledge and thrives on the kind of organizational cohesiveness found in the 'J-form' whereas exploration requires the creation of new knowledge and novel ideas nurtured in an entrepreneurial mode of organizing such as the adhocracy (Lam, 2000). The contrasting organizing logics underlying the two activities make their effective combination extremely difficult, if not impossible. However, in recent years there have been growing pressures on organizations to develop dual structures and processes for sustaining performance in a fast changing and complex environment. The notion of an 'ambidextrous organization' (O'Reilly and Tushman, 2004, 2008; Tushman *et al.,* 2010) suggests that the key to the long-term success of firms lies in their ability to exploit existing competences while simultaneously exploring new possibilities to compete in both mature and emerging markets. The term 'ambidexterity' means doing both. According to O'Reilly and Tushman (2004; 2008), ambidextrous organizations are ones that can sustain their competitive advantage by operating in multiple modes simultaneously—managing for short-term efficiency by emphasizing stability and control, and for long-term innovation by taking risks. Organizations that operate in this way develop multiple, internally inconsistent architectures, competences and cultures, with built-in capabilities for efficiency, consistency and reliability needed for exploiting current business on the one hand, and experimentation and improvisation for exploring new opportunities on the other. From a strategic perspective, organizational ambidexterity is seen as a dynamic capability enabling organizations "to maintain ecological fitness and, when necessary, to reconfigure existing assets and develop the new skills needed to address emerging threats and opportunities" (O'Reilly and Tushman, 2008: 189).

The concept of organizational ambidexterity is an attractive one. However, the conditions under which it leads to long-term success and its impact on innovative performance have yet to be verified. The challenge associated with managing the apparent paradox of stability and change remains a formidable task for many organizations.

5. CONCLUSION

Innovation is a process of learning, and learning is a collective process that occurs within an organized setting. This chapter has examined the nature and development of innovative organizations from three different but interdependent perspectives:
1. the relationship between organizational structural forms and innovativeness;
2. innovation as a process of organizational learning and knowledge creation; and
3. organizational capacity for change and adaptation. The analysis suggests that building innovative organizations entails not only matching structural forms with

technological and market opportunities, but also embedding the capacity for learning and knowledge creation within team processes and social relationships. There are different types of learning and innovative organizations and their dominant features tend to vary over time and across institutional contexts. However, a fundamental characteristic of innovation is that it always consists of a new combination of ideas, knowledge, capabilities and resources. Thus, maintaining the openness of an organization for absorbing new knowledge and ideas from a variety of sources increases the scope for new combinations and enhances the possibility for producing more sophisticated innovations. An enduring challenge facing all innovative organizations is the encapsulation of dual structures, processes and capabilities that reconcile stability and exploitation with change and exploration to ensure current viability and long-term adaptability. The notion of an 'ambidextrous organization' has become a popular expression to denote the paradox of managing innovation in the contemporary business environment.

Organizational innovation is a multifaceted phenomenon. The extensive literature in organization studies has advanced our understanding of the effects of organizational structure on the ability of organizations to learn, create knowledge and generate technological innovation. We know relatively less, however, about how internal organizational dynamics and actor learning interact with technological and environmental forces to shape organizational evolution. It remains unclear how and under what conditions organizations shift from one structural archetype to another, and the role of technological innovation in driving the process of organizational change is also obscure. The bulk of the existing research has tended to focus on how technology and market forces shape organizational outcomes and treat organizations primarily as a vehicle or facilitator of innovation, rather than focussing on the process of organizational innovation itself. For example, we tend to assume that technological innovation triggers organizational change because it shifts the competitive environment and forces organizations to adapt to the new set of demands. This deterministic view neglects the possibility that differences in organizational interpretations of, and responses to, external stimuli can affect the outcomes of organizational change. Treating the organization as an interpretation and learning system (e.g. Daft and Weick, 1994; Greve and Taylor, 2000) directs our attention to the important role of internal organizational dynamics, actor cognition and behaviour in shaping the external environment and outcomes of organizational change. A promising direction for future research would be to take greater account of endogenous organizational forces such as capacity for learning, values, interests and culture in shaping organizational change and innovation.

REFERENCES

AMABILE, T. M. (1988), "A model of creativity and innovation in organizations", in N. M. Staw and L. L. Cummings (eds.), *Research in Organizational Behaviour,* Greenwich, CT: JAI Press, 10, pp. 123-167.

ANGELS, D. P. (2000), "High-technology agglomeration and the labour market: the case of Silicon Valley", in K. Martin (ed.) *Understanding Silicon Valley: the Anatomy of an Entrepreneurial Region,* Stanford: Stanford University Press, pp. 125-189.

AOKI, M. (1988), *Information, Incentives and Bargaining in the Japanese Economy,* Cambridge: Cambridge University Press.

ARGOTE, L. (1999), *Organizational Learning: Creating, Retaining and Transferring Knowledge,* Norwell, MA: Kluwer Academic Publishers.

ARGYRIS, C., and D. SCHON (1978), *Organizational Learning: a Theory of Action Perspective,* Reading, MA: Addison-Wesley.

BAHRAMI, H. and S. EVANS (2000), "Flexible Recycling and High-Technology Entrepreneurship", in K. Martin (ed.), *Understanding Silicon Valley: the Anatomy of an Entrepreneurial Region,* Stanford: Stanford University Press, pp. 166-189.

BARTEL, C. A., and R. GARUD (2009), "The Role of Narratives in Sustaining Organizational Innovation", *Organization Science* 20, pp.107-117.

BARNETT, W. P. and G. R. CAROLL (1995), "Modelling Internal Organizational Change", *Annual Review of Sociology* 21, pp. 217-236.

BARNETT, W. P., and J. FREEMAN (2001), "Too Much of a Good thing? Product Proliferation and Organizational Failure", *Organization Science* 12, pp. 539-558.

BLAU, P. M. (1970), "A Formal Theory of Differentiation in Organizations", *American Sociological Review* 35, pp. 201-218.

BROWN, J. S., and P. DUGUID (1991), "Organizational Learning and Communities of Practice: Towards a Unified view of Working, Learning and Innovation," *Organization Science* 2, pp. 40-57.

BROWN, J. S. and P. DUGUID (1998), "Organizing Knowledge", *California Management Review* 40, pp. 90-111.

BROWN, S. L. and K. M. EISENHARDT (1997), "The Art of Continuous Change: Complexity Theory and Time-Paced Evolution in Relentlessly Shifting Organizations", *Administrative Science Quarterly* 42, pp. 1-34.

BURGLEMAN, R. A. (1991), "Intra-Organizational Ecology of Strategy-Making and Organizational Adaptation: Theory and Research", *Organization Science* 2, pp. 239-262.

BURGLEMAN, R. A. (2002), "Strategy as Vector and the Inertia of Co-evolutionary Lock-in", *Administrative Science Quarterly* 47, pp. 25-357.

BURNS,T., and G. M. STALKER (1961), *The Management of Innovation,* London: Tavistock.

CHANDLER, A. D. (1962), "Strategy and Structure: Chapters in the History of the American Industrial Enterprise", Cambridge, MA: The MIT Press.

CHILD, J. (1972), "Organizational Structure, Environment and Performance—the Role of Strategic Choice," *Sociology,* 6, pp. 1-22.

CHILD, J. (1997), Retrospect and Prospect", *Organization Studies* 18, pp. 43-76.

CHILD, J., and C. SMITH (1987), "The Context and Process of Organizational Transformation—Cadbury Limited in its Sector", *Journal of Management Studies* 24, pp. 565-593.

CHRISTENSEN, C. M. (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Boston, MA: Harvard Business School Press.

COHEN, W. M., and D. A. LEVINTHAL (1990), "Absorptive Capacity: a new Perspective on Learning and Innovation", *Administrative Science Quarterly* 35, pp.123-138.

DAFT, R. L., and K. E. WEICK (1984), "Toward a Model of Organizations as Interpretation Systems", *The Academy of Management Review* 9, pp. 284-295.

DEFILLIPPI, R. (2002), "Organization Models for Collaboration in the New Economy", *Human Resource Planning* 25, pp. 7-19.

DEFILLIPPI, R. J. and M. B. ARTHUR (1996), "Boundaryless Contexts and Careers: a Competency-Based Perspective" in M. B. Arthur and D. M. Rousseau (eds.), *The Boundaryless Career: A New Employment Principle for a New Organizational Era,* New York: Oxford University Press, pp. 116-131.

DIMAGGIO, P. J. and W. W. POWELL (1983), "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields," *American Sociological Review* 48, pp. 147-160.

DOSI, G. (1988), "Sources, Procedures, and Microeconomic Effects of Innovation", *Journal of Economic Literature* 26, pp. 1120-1171.

FARJOUN, M. (2010), "Beyond Dualism: Stability and Change as a Duality", *Academy of Management Review* 35, pp. 202-225.

FOSS, N. J. (2003) , "Selective Intervention and Internal Hybrids: Interpreting and Learning from the Rise and Decline of the Oticon Spaghetti Organization", *Organization Science* 14, pp. 331-349.

GALUNIC, D. C., and K. M. EISENHARDT (2001), "Architectural Innovation and Modular Corporate forms", *Academy of Management Journal* 44, pp. 1229-1249.

GLYNN, M. A. (1996), "Innovative Genius: a Framework for Relating Individual and Organizational Intelligence to Innovation", *Academy of Management Review* 21, pp.1081-1111.

GRANT, R. M. (1996), "Toward a Knowledge-Based Theory of the Firm", *Strategic Management Journal* 17, pp. 109-122.

GREENWOOD, R., and C. R. HININGS (1996), "Understanding Radical Organizational Change: Bringing Together the Old and New Institutionalism", *Academy of Management Review* 21, pp. 1022-1054.

GREVE, H. R. and A. TAYLOR (2000), "Innovations as Catalysts for Organizational Change: Shifts in Organizational Cognition and Change," *Administrative Science Quarterly* 45, pp. 54-80.

HANNAN, M. T. and J. H. FREEMAN (1984), "Structural Inertia and Organizational Change", *American Sociological Review* 49: pp. 149-164.

HEDLUND, G. (1994), "A Model of Knowledge Management and the n-Form Corporation", *Strategic Management Journal* 15: pp. 73-90.

KANTER, R. M. (1983), *The Change Masters,* New York: Simon & Schuster.

KOGUT, B., and U. ZANDER (1992), "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology", *Organization Science* 3, pp. 383-397.

LAM, A. (2000), "Tacit knowledge, Organizational Learning, Societal Institutions: an Integrated Framework", *Organization Studies* 21, pp. 487-513.

LAM, A. (2002), "Alternative Societal Models of Learning and Innovation in the Knowledge Economy", *International Social Science Journal* 17, pp. 67-82.

LAM, A. (2007), "Knowledge Networks and Careers: Academic Scientists in Industry-University Links", *Journal of Management Studies* 44, pp. 993-1016.

LAM, A. and B-Å. LUNDVALL (2006), "The Learning Organization and National Systems of Competence Building and Innovation" in E. Lorenz and B- Å. Lundvall (eds.), *How Europe's Economies Learn: Coordinating Competing Models,* Oxford: Oxford University Press, pp. 109-139.

LAVE, J., and E. WENGER (1991), *Situated Learning: Legitimate Peripheral Participation,* New York: Cambridge University Press.

LAWRENCE, P. R., and J. W. LORSCH (1967), "Differentiation and Integration in Complex Organizations", *Administrative Science Quarterly* 12, pp. 1-47.

LAZONICK, W. (2004), "The Innovative Firm," in J. Fagerberg, D. Mowery and R. Nelson (eds.), *The Oxford Handbook of Innovation,* Oxford: Oxford University Press, pp. 29-55.

LAZONICK, W. (2010), "The Chandlerian Corporation and the Theory of Innovative Enterprise", *Industrial and Corporate Change* 1, pp. 317-349.

LEONARD-BARTON, D. (1992), "Core Capabilities and Core Rigidities: a Paradox in Managing New Product Development", *Strategic Management Journal* 13, pp. 111-125.

LEVINTHAL, D. A., and J. G. MARCH (1993), "The Myopia of Learning", *Strategic Management Journal* 14, pp. 95-112.

LEWIN, A. Y., and H. W. VOLBERDA (1999), "Prolegomena on Co-evolution: a Framework for Research on Strategy and New Organizational Forms", *Organization Science* 10, pp. 519-534.

LUNDVALL, B.-Å. (ed.) (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London: Pinter.

March, J. G. (1991), "Exploration and Exploitation in Organizational Learning", *Organization Science* 2, pp.71-87.

Mezias, S. J., and M. A. Glynn (1993), "The Three Faces of Corporate Renewal: Institution, Revolution, and Evolution", *Strategic Management Journal* 14: pp. 77-101.

Miles, R. E., C. C. Snow, J. A. Mathews, G. Miles and H. J. Coleman Jr. (1997), "Organizing in the Knowledge Age: Anticipating the Cellular form", *Academy of Management Executive* 11(4), pp. 7-20.

Mintzberg, H. (1979), *The Structuring of Organization,* Englewood Cliffs, NJ: Prentice Hall.

Nelson, R. R., and S.G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: The Belknap Press of Harvard University Press.

Nonaka., I., and H. Takeuchi (1995), *The Knowledge Creating Company,* New York: Oxford University Press.

Nonaka, I. (1994), "A dynamic theory of organizational knowledge creation", *Organization Science* 5, pp. 14-37.

Nonaka, I., and G. von Krogh (2009), "Perspective—Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory", *Organization Science* 20, pp. 635-652.

O'Reilly, C. A., and M. L. Tushman (2004), "The Ambidextrous Organization", *Harvard Business Review,* April, pp. 74-81.

O'Reilly, C. A. and M. L. Tushman (2008), "Ambidexterity as a Dynamic Capability: Resolving the Innovator's Dilemmas", *Research in Organizational Behaviour* 28, pp. 185-206.

Pavitt, K. (1991), "Key Characteristics of the Large Innovating Firm", *British Journal of Management* 2, pp. 41-50.

Perrow, C. 1970, *Organizational Analysis,* London: Tavistock.

Polanyi, M. (1966), *The Tacit Dimension,* New York: Anchor Day Books.

Powell, W. W. (1998), "Learning from Collaboration: Knowledge and Networks in the Biotechnology and Pharmaceutical Industries", *California Management Review* 40, pp. 228-240.

Prahalad, C. K. and G. Hamel (1990), "The Core Competence of the Corporation", *Harvard Business Review,* May/June: pp.79-91.

Quinn, J. B. (1992), *Intelligent Enterprise: a Knowledge and Service Based Paradigm for Industry,* New York: The Free Press.

Saxenian, A. (1996), "Beyond Boundaries: Open Labour Markets and Learning in the Silicon Valley" in M. B. Arthur and D. M. Rousseau (eds.), *The Boundaryless Career: a New Employment Principle for a New Organizational Era,* New York: Oxford University Press, pp. 23-39.

Schumpeter, J. (1950), "The Process of Creative Destruction," in J. Schumpeter (ed.) *Capitalism, Socialism and Democracy,* 3rd ed., London: Allen and Unwin.

Senge, P. (1990), *The Fifth Discipline: the Art and Practice of the Learning Organization* New York: Doubleday.

Simon, H. A. (1991), "Bounded Rationality and Organizational Learning", *Organization Science* 2, pp.125-134.

Teece, D. J., G. Pisano and A. Shuen (1997), "Dynamic Capabilities and Strategic Management", *Strategic Management Journal* 18, pp. 509-533.

Teece, D. (2007), "Explicating Dynamic Capabilities: the Nature and Microfoundations of (Sustainable) Enterprise Performance", *Strategic Management Journal* 28, pp. 1319-1350.

Tidd, J., J. Bessant and K. Pavitt (1997), *Managing Innovation,* Chichester: John Wiley &Sons.

Tushamn, M. L., W. K. Smith and R. C. Wood *et al.* (2010), "Organizational Designs and Innovation Streams", *Industrial and Corporate Change,* doi: 10.1093/icc/dtq040.

Van de Ven, A., D. Polley, S. Garud and S. Venkataraman (1999), *The Innovation Journey,* New York: Oxford University Press.

Von Hippel, E. (1988), *The Sources of Innovation,* New York: Oxford University Press.

Walsh, J. P., and G. R. Ungson (1991), "Organizational Memory", *Academy of Management Review* 16, 57-91.

Weber, M. (1947), *The Theory of Social and Economic Organization,* Glencoe, IL: The Free Press.

Weick, K. E. (1979), *The Social Psychology of Organizing,* 2nd ed., Reading, MA: Addison-Wesley.

Weick, K. E. (1995), *Sense-making in Organizations,* Thousand Oaks, CA: Sage.

Weick, K. E. (1996), "The Role of Renewal in Organizational Learning", *International Journal of Technology Management* 11, pp. 738-746.

Wenger, E. (1998), *Communities of Practice: Learning, Meaning, and Identity*. New York, Cambridge University Press.

Whitley, R. (2000), "The Institutional Structuring of Innovation Strategies: Business Systems, Firm Types and Patterns of Technical Change in Different Market Economies", *Organization Studies* 21, pp. 855-886.

Whitley, R. (2003), "The Institutional Structuring of Organizational Capabilities: the Role of Authority Sharing and Organizational Careers", *Organization Studies* 24, pp. 667-695.

Womack, J. P., D. T. Jones and D. Roos (1990), *The Machine that Changed the World.* New York: Rawson Associates.

Woodman, R. W., J. E. Sawyer and R. W. Griffin (1993), "Toward a Theory of Organizational Creativity", *Academy of Management Review* 18, pp. 293-321.

Woodward, J. (1965), *Industrial Organization, Theory and Practice,* London: Oxford University Press.

BBVA

Innovation by Users

Eric von Hippel
Massachusetts Institute of Technology (MIT)

Ever since Joseph A. Schumpeter (1934) promulgated his theory of economic development, economists, policymakers and business managers have assumed that the dominant mode of innovation is a "producers' model." That is, it has been assumed that most important innovations would originate from producers and be supplied to consumers via goods that were for sale.

This view seemed reasonable on the face of it—producers generally serve many users and so can profit from multiple copies of a single innovative design. Individual users, in contrast, depend upon benefits from in-house use of an innovation to recoup their investments. Presumably, therefore, a producer that serves many customers can afford to invest more in innovation than any single user. From this it follows logically that producer-developed designs should dominate user-developed designs in most parts of the economy.

However, the producers' model is only one mode of innovation. A second, increasingly important model is *user innovation*. Under this second model, economically important innovations are developed by individual users (consumers) and also by user firms. Sometimes, user-developed innovations result from a number of users working together collaboratively.

User innovation is an institution that competes with and, I will argue, can displace producer innovation in many parts of the economy. A growing body of empirical work clearly shows that users are the first to develop many and perhaps most new industrial and consumer products. In addition, the importance of product and service development by users is increasing over time. This shift is being driven by two related technical trends: 1. the steadily improving *design capabilities* (innovation toolkits) that advances in computer hardware and software make possible for users; 2. the steadily improving ability of individual users to *combine and coordinate* their innovation-related efforts via new communication media such as the Internet.

The ongoing shift of innovation to users has some very attractive qualities. It is becoming progressively easier for many users to get precisely what they want by designing it for themselves. Innovation by users also provides a very necessary complement to, and feedstock for, producer

> "The ongoing shift of innovation to users has some very attractive qualities. It is becoming progressively easier for many users to get precisely what they want by designing it for themselves. Innovation by users also provides a very necessary complement to, and feedstock for, producer innovation"

innovation. And innovation by users appears to increase social welfare. At the same time, the ongoing shift of product-development activities from producers to users is painful and difficult for many producers. User innovation is "attacking" a major structure of the social division of labor. Many firms and industries must make fundamental changes to long-held business models in order to adapt. Further, governmental policy and legislation sometimes preferentially supports innovation by producers. Considerations of social welfare suggest that this must change. The workings of the intellectual property system are of special concern. But despite the difficulties, a user-centered system of innovation appears well worth striving for.

Today a number of innovation process researchers are working to develop our understanding of user innovation processes. In this paper, I offer a review of some collective learning on this important topic to date.

IMPORTANCE OF INNOVATION BY USERS

Users, as I use the term, are firms or individual consumers that expect to benefit from *using* a product or a service. In contrast, producers expect to benefit from *selling* a product or a service. A firm or an individual can have different relationships to different products or innovations. For example, Boeing is a producer of airplanes, but it is also a user of machine tools. If one were examining innovations developed by Boeing for the airplanes it sells, Boeing would be a producer-innovator in those cases. But if one were considering innovations in metal-forming machinery developed by Boeing for in-house use in building airplanes, those would be categorized as user-developed innovations and Boeing would be a user-innovator in those cases.

Innovation user and innovation producer are the two general "functional" relationships between innovator and innovation. Users are unique in that they alone benefit *directly* from innovations. All others (here lumped under the term "producers") must sell innovation-related products or services to users, indirectly or directly, in order to profit from innovations. Thus, in order to profit, inventors must sell or license knowledge related to innovations, and producers must sell products or services incorporating innovations. Similarly, suppliers of innovation-related materials or services—unless they have direct use for the innovations—must sell the materials or services in order to profit from the innovations.

The user and producer categorization of relationships between innovator and

> “Today a number of innovation process researchers are working to develop our understanding of user innovation processes”

innovation can be extended to specific function, attributes, or features of products and services. When this is done, it may turn out that different parties are associated with different attributes of a particular product or service. For example, householders are the users of the switching attribute of a household electric light switch—they use it to turn lights on and off. However, switches also have other attributes, such as “easy wiring” qualities, that may be used only by the electricians who install them. Therefore, if an electrician were to develop an improvement to the installation attributes of a switch, it would be considered a user-developed innovation.

Both qualitative observations and quantitative research in a number of fields clearly document the important role users play as first developers of products and services later sold by manufacturing firms. Adam Smith (1776) was an early observer of the phenomenon, pointing out the importance of “the invention of a great number of machines which facilitate and abridge labor, and enable one man to do the work of many”. Smith went on to note that “a great part of the machines made use of in those manufactures in which labor is most subdivided, were originally the invention of common workmen, who, being each of them employed in some very simple operation, naturally turned their thoughts towards finding out easier and readier methods of performing it”. Rosenberg (1976) explored the matter in terms of innovation by *user firms* rather than individual workers. He studied the history of the US machine tool industry, finding that important and basic machine types like lathes and milling machines were first developed and built by user firms having a strong need for them. Textile manufacturing firms, gun producers and sewing machine producers were important early user-developers of machine tools.

Quantitative studies of user innovation document that many of the most important and novel products and processes in a range of fields have been developed by user firms and by individual users. Thus, Enos (1962) reported that nearly all the most important innovations in oil refining were developed by user firms. Freeman (1968) found that the most widely licensed chemical production processes were developed by user firms. Von Hippel (1988) found that users were the developers of about 80% of the most important scientific instrument innovations, and also the developers of most of the major innovations in semiconductor processing. Pavitt (1984) found that a considerable fraction of invention by British firms was for in-house use. Shah (2000) found that the most commercially important equipment innovations in four sporting fields tended to be developed by individual users.

Empirical studies also show that *many* users—from 10% to nearly 40%—engage in developing or modifying products. This has been documented in the case of specific types of industrial products and consumer products, and in large, multi-industry studies of process innovation in Canada and the Netherlands as well (table 1). When taken together, the findings make it very clear that users are doing a *lot* of product development and product modification in many fields.

Studies of innovating users (both individuals and firms) show them to have the characteristics of "lead users" (Urban and von Hippel, 1988, Herstatt and von Hippel, 1992, Olson and Bakke, 2001, Lilien *et al.,* 2002). That is, they are ahead of the majority of users in their populations with respect to an important market trend, and they expect to gain relatively high benefits from a solution of the needs they have encountered there. The correlations found between innovation by users and lead user status are highly significant, and the effects are considerable (Franke and Shah, 2003, Lüthje *et al.,* 2002 and Morrison *et al.,* 2000).

Since lead users are at the leading edge of the market with respect to important market trends, one can guess that many of the novel

Table 1. Studies of user innovation frequency

Innovation Area	Number and type of users sampled	% developing and building product for own use
Industrial products		
1. Printed Circuit CAD Software (a)	136 user firm attendees at a PC-CAD conference	24.3%
2. Pipe Hanger Hardware (b)	Employees in 74 pipe hanger installation firms	36%
3. Library Information Systems (c)	Employees in 102 Australian libraries using computerized OPAC library information systems	26%
4. Medical Surgery Equipment (d)	261 surgeons working in university clinics in Germany	22%
5. Apache OS server software security features (e)	131 technically sophisticated Apache users (webmasters)	19.1%
Consumer products		
6. Outdoor consumer products (f)	153 recipients of mail order catalogs for outdoor activity products for consumers	9.8%
7. "Extreme" sporting equipment (g)	197 members of 4 specialized sporting clubs in 4 "extreme" sports	37.8%
8. Mountain biking equipment (h)	291 mountain bikers in a geographic region known to be an "innovation hot spot."	19.2%
Multi-industry process innovation surveys		
26. 'Advanced Manufacturing Technologies' (i)	Canadian manufacturing plants in 9 Manufacturing Sectors (less food processing) in Canada, 1998 (population estimates based upon a sample of 4,200)	28% developed 26% modified
39. 'Advanced Manufacturing Technologies' (j)	16,590 Canadian manufacturing establishments that met the criteria of having at least $250,000 in revenues, and at least 20 employees.	22% developed 21% modified
Any type of process innovation or process modification (k)	Representative, cross-industry sample of 498 "high tech" Netherlands SMEs	41% developed only 34% modified only 54% developed and/or modified

Data Sources: a. Urban and von Hippel, 1988; b. Herstatt and von Hippel,1992; c. Morrison *et al.,* 2000; d. Lüthje, 2003; e. Franke and von Hippel, 2003; f. Lüthje, 2004; g. Franke and Shah, 2003; h. Lüthje *et al.,* 2002; i. Arundel and Sonntag, 1999; j. Gault and von Hippel, 2009; k. de Jong and von Hippel, 2009.

products they develop for their own use will appeal to other users too and so might provide the basis for products producers would wish to commercialize. This turns out to be the case. A number of studies have shown that many of the innovations reported by lead users are judged to be commercially attractive and/or have actually been commercialized by producers.

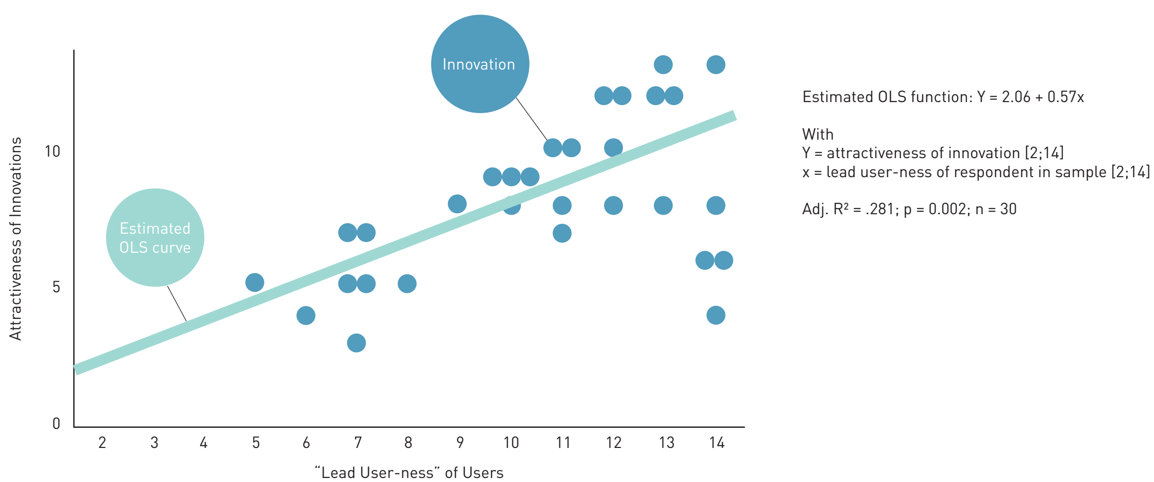
Research provides a firm grounding for these empirical findings. The two defining characteristics of lead users and the likelihood that they will develop new or modified products have been found to be highly correlated (Morrison *et al.,* 2004). In addition, it has been found that the higher the intensity of lead user characteristics displayed by an innovator, the greater the commercial attractiveness of the innovation that that lead user develops (Franke and von Hippel, 2003a). In figure 1, the increased concentration of innovations toward the right indicates that the likelihood of innovating is higher for users having higher lead user index values. The rise in average innovation attractiveness as one moves from left to right indicates that innovations developed by lead users tend to be more commercially attractive. (Innovation attractiveness is the sum of the novelty of the innovation and the expected future generality of market demand.)

WHY MANY USERS WANT CUSTOM PRODUCTS

Why do so many users develop or modify products for their own use? Users may innovate if and as they want something that is not available on the market and are able and willing to pay for its development. It is likely that many users do not find what they want on the market. Meta-analysis of market-segmentation studies suggests that users' needs for products are highly heterogeneous in many fields (Franke and Reisinger, 2003).

Mass producers tend to follow a strategy of developing products that are designed to meet the needs of a large market segment well enough to induce purchase from and capture significant profits from a large number of customers. When users' needs

Figure 1. User-innovators with stronger "lead user" characteristics develop innovations having higher appeal in the general market place



Data Source: Franke and von Hippel, 2003 are heterogeneous, this strategy of "a few sizes fit all" will leave many users somewhat dissatisfied with the commercial products on offer and probably will leave some users seriously dissatisfied. In a study of a sample of users of the security features of Apache web server software, Franke and von Hippel (2003b) found that users had a very high heterogeneity of need, and that many had a high willingness to pay to get precisely what they wanted. 19% of the users sampled actually innovated to tailor Apache more closely to their needs. Those who did were found to be significantly more satisfied.

> "Mass producers tend to follow a strategy of developing products that are designed to meet the needs of a large market segment well enough to induce purchase from and capture significant profits from a large number of customers"

USERS' INNOVATE-OR-BUY DECISIONS

Even if many users want "exactly-right products" and are willing and able to pay for their development, we must understand why users often do this for themselves rather than hire a custom producer to develop a special just-right product for them. After all, custom producers specialize in developing products for one or a few users. Since these firms are specialists, it is possible that they could design and build custom products for individual users or user firms faster, better, or cheaper than users could do this for themselves. Despite this possibility, several factors can drive users to innovate rather than buy. Both in the case of user firms and that of individual user-innovators, agency costs play a major role. In the case of individual user-innovators, enjoyment of the innovation process can also be important.

With respect to agency costs, consider that when a user develops its own custom product that user can be trusted to act in its own best interests. When a user hires a producer to develop a custom product, the situation is more complex. The user is then a principal that has hired the custom producer to act as its agent. If the interests of the principal and the agent are not the same, there will be agency costs. In general terms, agency costs are 1. costs incurred to monitor the agent to ensure that it (or he or she) follows the interests of the principal, 2. the cost incurred by the agent to commit itself not to act against the principal's interest (the "bonding cost"), and 3. costs associated with an outcome that does not fully serve the interests of the principal (Jensen and Meckling, 1976). In the specific instance of product and service development, a major divergence of interests between user and custom producer does exist: the user wants to get precisely what it needs, to the extent that it can afford to do so. In contrast, the custom producer wants to lower its development costs by incorporating solution elements it already has or that it predicts others will want in the future—even if by doing so it does not serve its present client's needs as well as it could.

A user wants to preserve its need specification because that specification is

> “The custom producer wants to lower its development costs by incorporating solution elements it already has or that it predicts others will want in the future—even if by doing so it does not serve its present client’s needs as well as it could”

chosen to make *that user’s* overall solution quality as high as possible at the desired price. For example, an individual user may specify a mountain-climbing boot that will precisely fit his unique climbing technique and allow him to climb Everest more easily. Any deviations in boot design will require compensating modifications in the climber’s carefully practiced and deeply ingrained climbing technique—a much more costly solution from the user’s point of view. A custom boot producer, in contrast, will have a strong incentive to incorporate the materials and processes it has in stock and expects to use in future even if this produces a boot that is not precisely right for the present customer. For example, the producer will not want to learn a new way to bond boot components together even if that would produce the best custom result for one client. The net result is that when one or a few users want something special they will often get the best result by innovating for themselves.

A model of the innovate-or-buy decision (von Hippel, 2005) shows in a quantitative way that user firms with unique needs (in other words, a market of one) will always be better off developing new products for themselves. It also shows that development by producers can be the most economical option when n or more user firms want the same thing. However, when the number of user firms wanting the same thing is between 1 and n, producers may not find it profitable to develop a new product for just a few users. In that case, more than one user may invest in developing the same thing independently, owing to market failure. This results in a waste of resources from the point of view of social welfare. The problem can be addressed by new institutional forms, such as the user innovation communities that will be mentioned later.

It is important to note that an additional incentive can drive individual user-innovators to innovate rather than buy: they may value the *process* of innovating because of the enjoyment or learning that it brings them. It might seem strange that user-innovators can enjoy product development enough to want to do it themselves—after all, producers pay their product developers to do such work! On the other hand, it is also clear that enjoyment of problem solving is a motivator for many individual problem solvers in at least some fields. Consider for example the millions of crossword-puzzle *aficionados*. Clearly, for these individuals enjoyment of the problem-solving process rather than the solution is the goal. One can easily test this by attempting to offer a puzzle solver a completed puzzle—the very output he or she is working so hard to create. One will very likely be rejected with the rebuke that one should not spoil the fun. Pleasure as a motivator can apply to the development of commercially useful innovations as well. Studies of the motivations of volunteer contributors of code to widely used software products have shown that these individuals too are often strongly motivated to innovate by the joy and learning they find in this work (Hertel *et al.,* 2003; Lakhani and Wolf, 2005).

USERS' LOW-COST INNOVATION NICHES

An exploration of the basic processes of product and service development shows that users and producers tend to develop different *types* of innovations. This is due in part to information asymmetries: users and producers tend to know different things. Product developers need two types of information in order to succeed at their work: need and context-of-use information (generated by users) and generic solution information (often initially generated by producers specializing in a particular type of solution). Bringing these two types of information together is not easy. Both need information and solution information are often very "sticky"—that is, costly to move from the site where the information was generated to other sites (von Hippel, 1994). It should be noted that the observation that information is often sticky contravenes a central tendency in economic theorizing. Much of the research on the special character of markets for information, and the difficulty of appropriating benefit from invention and innovation, has been based on the idea that information can be transferred at very low cost. (Thus, Arrow observes that "the cost of transmitting a given body of information is frequently very low. . . In the absence of special legal protection, the owner cannot, however, simply sell information on the open market. Any one purchaser can destroy the monopoly, since he can reproduce the information at little or no cost" (1962: 614-615).

> "Product developers need two types of information in order to succeed at their work: need and context-of-use information (generated by users) and generic solution information (often initially generated by producers specializing in a particular type of solution)"

When information is sticky, innovators tend to rely largely on information they already have in stock. One consequence of the resulting typical asymmetry between users and producers is that users tend to develop innovations that are functionally novel, requiring a great deal of user-need information and use-context information for their development. In contrast, producers tend to develop innovations that are improvements on well-known needs and that require a rich understanding of solution information for their development. Similarly, users tend to have better information regarding ways to improve use-related activities such as maintenance than do producers: they "learn by using" (Rosenberg, 1982).

This sticky information effect is quantitatively visible in studies of innovation.

Riggs and von Hippel (1994) studied the types of innovations made by users and producers that improved the functioning of two major types of scientific instruments. They found that users are significantly more likely than producers to develop innovations that enabled the instruments to do qualitatively new types of things for the first time. In contrast, producers tended to develop innovations that enabled users to do the same things they had been doing, but to do them more conveniently or reliably (table 2). For example, users were the first to modify the instruments to enable them to image and analyze magnetic domains at sub-microscopic dimensions. In contrast, producers were the first to computerize instrument adjustments to improve ease of operation. Sensitivity, resolution, and accuracy improvements fall somewhere in the middle, as the data show. These types of improvements can be driven by users seeking to do specific new things, or by producers applying their technical expertise to improve the products along known general dimensions of merit, such as accuracy.

The sticky information effect is independent of Stigler's (1951) argument that the division of labor is limited by the extent of the market. When profit expectations are controlled, the impact of sticky information on the locus of innovation is still strongly evident (Ogawa, 1998).

If we extend the information-asymmetry argument one step further, we see that information stickiness implies that information on hand will also differ among *individual* users and producers. The information assets of some particular user (or some particular producer) will be closest to what is required to develop a particular innovation, and so the cost of developing that innovation will be relatively low for that user or producer. The net result is that user innovation activities will be *distributed* across many users according to their information endowments. With respect to innovation, one user is by no means a perfect substitute for another.

WHY USERS OFTEN FREELY REVEAL THEIR INNOVATIONS

The social efficiency of a system in which individual innovations are developed by individual users is increased if users somehow pass on what they have developed to others. Producer-innovators *partially* achieve this when they sell a product or a service on the open market (partially because they disseminate the product incorporating the innovation, but often not all the information that others would need to fully understand and replicate it). If user-innovators do not somehow also pass on what they have done, multiple users with

Table 2. Source of innovations by nature of improvement effected

Type of improvement provided by innovation	Innovation developed by:			
	%User	User	Producer	Total
1. New functional capability	82%	14	3	17
2. Sensitivity, resolution or accuracy improvement	48%	11	12	23
3. Convenience or reliability improvement	13%	3	21	24
Total				**64**

Source: Riggs and von Hippel, 1994.

> “Hiding an innovation as a trade secret is unlikely to be successful for long: too many generally know similar things, and some holders of the “secret” information stand to lose little or nothing by freely revealing what they know”

very similar needs will have to independently develop very similar innovations—a poor use of resources from the viewpoint of social welfare. Empirical research shows that users often do achieve widespread diffusion by an unexpected means: they often “freely reveal” what they have developed. When we say that an innovator freely reveals information about a product or service it has developed, we mean that all intellectual property rights to that information are voluntarily given up by the innovator, and all interested parties are given access to it—the information becomes a public good (Harhoff *et al.,* 2003).

The empirical finding that users often freely reveal their innovations has been a major surprise to innovation researchers. On the face of it, if a user-innovator’s proprietary information has value to others, one would think that the user would strive to prevent free diffusion rather than help others to a free ride on what it has developed at private cost. Nonetheless, it is now very clear that individual users and user firms—and sometimes producers—often freely reveal detailed information about their innovations.

The practices visible in “open source” software development were important in bringing this phenomenon to general awareness. In these projects it was clear *policy* that project contributors would routinely and systematically freely reveal code they had developed at private expense (Raymond, 1999). However, free revealing of product innovations has a history that began long before the advent of open source software. Allen, in his 1983 study of the eighteenth-century iron industry, was probably the first to consider the phenomenon systematically. Later, Nuvolari (2004) discussed free revealing in the early history of mine pumping engines. Contemporary free revealing by users has been documented by von Hippel and Finkelstein (1979) for medical equipment, by Lim (2000) for semiconductor process equipment, by Morrison, Roberts, and von Hippel (2000) for library information systems, and by Franke and Shah (2003) for sporting equipment. Henkel (2003) has documented free revealing among producers in the case of embedded Linux software.

Innovators often freely reveal because it is often the best or the only practical option available to them. Hiding an innovation as a trade secret is unlikely to be successful for long: too many generally know similar things, and some holders of the “secret” information stand to lose little or nothing by freely revealing what they know. Studies find that innovators in many fields view patents as having only limited value (Harhoff *et al.,* 2003). Copyright protection and copyright licensing are applicable only to “writings,” such as books, graphic images, and computer software.

Active efforts by innovators to freely reveal—as opposed to sullen acceptance—are explicable because free revealing can provide innovators with significant private benefits as well as losses or risks of loss. Users who freely reveal what they have done often find that others then improve or suggest improvements to the innovation, to their mutual benefit (Raymond, 1999). Freely- revealing users also may benefit from enhancement of reputation, from positive network effects due to increased diffusion of their innovation, and from other factors. Being the first to freely reveal a particular innovation can also enhance the benefits received, and so there can actually be a rush to reveal, much as scientists rush to publish in order to gain the benefits associated with being the first to have made a particular advance.

INNOVATION COMMUNITIES

Innovation by users tends to be widely distributed rather than concentrated among just a very few very innovative users (table 3). As a result, it is important for user-innovators to find ways to combine and leverage their efforts. Users achieve this by engaging in many forms of cooperation. Direct, informal user-to-user cooperation (assisting others to innovate, answering questions, and so on) is common. Organized cooperation is also common, with users joining together in networks and communities that provide useful structures and tools for their interactions and for the distribution of innovations. Innovation communities can increase the speed and effectiveness with which users and also producers can develop, test and diffuse their innovations. They also can greatly increase the ease with which innovators can build larger systems from interlinkable modules created by community participants.

Free and open source software projects are a relatively well-developed and very successful form of an Internet-based innovation community. However, innovation communities are by no means restricted to software or even to information products, and they can play a major role in the development of physical products. Franke and Shah (2003) have documented the value that user-innovation communities can provide to user-innovators developing physical products in the field of sporting equipment. The analogy to open source innovation communities is clear.

The collective or community effort to provide a public good—which is what freely revealed innovations are—has traditionally been explored in the literature on "collective action". However, behaviors seen in extant innovation communities fail to correspond to

Table Source: von Hippel, 2005, table 7-1.

Data Sources:
* von Hippel, 1988, Appendix: GC, TEM, NMR Innovations
** Riggs and von Hippel, Esca and AES
*** von Hippel, 1988, Appendix: Semiconductor and pultrusion process equipment innovations.
**** Shah, 2000, Appendix A: skateboarding, snowboarding and windsurfing innovations developed by users.

Table 3. User innovation is widely distributed: Few users developed more than one major commercialized innovation

User samples	Number of innovations each user developed:					
	1	2	3	6	na	sample (n)
Scientific Instrument users*	28	0	1	0	1	32
Scientific Instrument users**	20	1	0	1	0	28
Process equipment users***	19	1	0	0	8	29
Sports equipment users****	7	0	0	0	0	7

that literature at major points. In essence, innovation communities appear to be more robust with respect to recruiting and rewarding members than the literature would predict. The reason for this appears to be that innovation contributors obtain some private rewards that are not shared equally by free riders (those who take without contributing). For example, a product that a user-innovator develops and freely reveals might be perfectly suited to that user-innovator's requirements but less well suited to the requirements of free riders. Innovation communities thus illustrate a "private-collective" model of innovation incentive (von Hippel and von Krogh, 2003).

> "This finding implies that policy making should support user innovation, or at least should ensure that legislation and regulations do not favor producers at the expense of user-innovators"

ADAPTING POLICY TO USER INNOVATION

Is innovation by users a "good thing"? Welfare economists answer such a question by studying how a phenomenon or a change affects social welfare. Henkel and von Hippel (2005) explored the social welfare implications of user innovation. They found that, relative to a world in which only producers innovate, social welfare is very probably increased by the presence of innovations freely revealed by users. This finding implies that policy making should support user innovation, or at least should ensure that legislation and regulations do not favor producers at the expense of user-innovators.

The transitions required of policy making to achieve neutrality with respect to user innovation vs. producer innovation are significant. Consider the impact on open and distributed innovation of past and current policy decisions. Research done in the past 30 years has convinced many academics that intellectual property law is sometimes or often not having its intended effect. Intellectual property law was intended to increase the amount of innovation investment. Instead, it now appears that there are economies of scope in both patenting and copyright that allow firms to use these forms of intellectual property law in ways that are directly opposed to the intent of policy makers and to the public welfare (Foray, 2004). Major firms can invest to develop large portfolios of patents. They can then use these to create "patent thickets"—dense networks of patent claims that give them plausible grounds for threatening to sue across a wide range of intellectual property. They may do this to prevent others from introducing a superior innovation and/or to demand licenses from weaker competitors on favorable terms (Shapiro, 2001; Bessen, 2003). Movie, publishing, and software firms can use large collections of copyrighted work for a similar purpose (Benkler, 2002). In view of the distributed nature of innovation by users, with each tending to create a relatively small amount of intellectual property, users are likely to be disadvantaged by such strategies.

It is also important to note that users (and producers) tend to build prototypes of their innovations economically by modifying products already available on the market to serve a new purpose. Laws such as the (US) Digital Millennium Copyright Act, intended to prevent consumers from illegally copying protected works, also can have the unintended side effect of preventing users from modifying products that they purchase (Varian, 2002). Both fairness and social welfare considerations suggest that innovation-related policies should be made neutral with respect to the sources of innovation.

It may be that current impediments to user innovation will be solved by legislation or by policy making. However, beneficiaries of existing law and policy will predictably resist change. Fortunately, a way to get around some of these problems is in the hands of innovators themselves. Suppose many innovators in a particular field decide to freely reveal what they have developed, as they often have reason to do. In that case, users can collectively create an information commons (a collection of information freely available to all) containing substitutes for some or a great deal of information now held as private intellectual property. Then user-innovators can work around the strictures of intellectual property law by simply using these freely revealed substitutes (Lessig, 2001).

This pattern is occurring in the field of software—and very visibly so. For many problems, user-innovators in that field now have a choice between proprietary, closed software provided by Microsoft and other firms and open-source software that they can legally download from the Internet and legally modify as they wish, to serve their own specific needs. It is also happening, although less visibly, in the case of process equipment developed by users for in-house use. Data from both Canada and the Netherlands show that about 25% of such user-developed innovations get voluntarily transferred to producers. A significant fraction—about half—is transferred both unprotected by intellectual property and without charge (Gault and von Hippel, 2009, de Jong and von Hippel, 2009).

Policy making that levels the playing field between users and producers will force more rapid change onto producers but will by no means destroy them. Experience in fields where open and distributed innovation processes are far advanced show how producers can and do adapt. Some, for example, learn to supply proprietary platform products that offer user-innovators a framework upon which to develop and use their improvements (Jeppesen, 2004).

DIFFUSION OF USER-DEVELOPED INNOVATIONS

Products, services, and processes developed by users become more valuable to society if they are somehow diffused to others that can also benefit from them. If user innovations are not diffused, multiple users with very similar needs will have to invest to (re)develop very similar innovations which, as was noted earlier, would be a poor use of resources from the social-welfare point of view. In the case of information products, users have the possibility of largely or completely doing without the services of producers. Open-source software projects are object lessons that teach us that users can create, produce, diffuse and provide user field support, update, and use complex products by and for themselves in the context of user innovation communities. In physical product fields, the situation is different. Users can develop products. However, the economies of scale associated with manufacturing and distributing physical products give producers

> "To systematically find user-developed innovations, producers must redesign their product development processes"

an advantage over "do-it-yourself" users in those activities.

How can or should user innovations of general interest be transferred to producers for large-scale diffusion? We propose three general methods for accomplishing this. First, producers can actively seek innovations developed by lead users that can form the basis for a profitable commercial product. Second, producers can draw innovating users into joint design interactions by providing them with "toolkits for user innovation." Third, users can become producers in order to widely diffuse their innovations. We discuss each of these possibilities in turn.

To systematically find user-developed innovations, producers must redesign their product development processes. Currently, almost all producers think that their job is to find a need and fill it rather than to sometimes find and commercialize an innovation that lead users have already developed. Accordingly, producers have set up market-research departments to explore the needs of users in the target market, product-development groups to think up suitable products to address those needs, and so forth. In this type of product development system, the needs and prototype solutions of lead users—if encountered at all—are typically rejected as outliers of no interest. Indeed, when lead users' innovations do enter a firm's product line they typically arrive with a lag and by an unconventional and unsystematic route. For example, a producer may "discover" a lead user innovation only when the innovating user firm contacts the producer with a proposal to produce its design in volume to supply its own in-house needs. Or sales or service people employed by a producer may spot a promising prototype during a visit to a customer's site.

Modification of firms' innovation processes to *systematically* search for and further develop innovations created by lead users can provide producers with a better interface to the innovation process as it actually works, and so provide better performance. A natural experiment conducted at 3M illustrates this possibility. Annual sales of lead user product ideas generated by the average lead user project at 3M were conservatively forecast by management to be more than 8 times the sales forecast for new products developed in the traditional manner—$146 million versus $18 million per year. In addition, lead user projects were found to generate ideas for new product lines, whereas traditional market-research methods were only found to produce ideas for incremental improvements to existing product lines. As a consequence, 3M divisions funding lead user project ideas experienced their highest rate of major product line generation in the past 50 years (Lilien *et al.,* 2002).

Toolkits for user innovation custom design involve partitioning product-development and service-development projects into solution-information-intensive subtasks and need-information-intensive subtasks. Need-intensive subtasks are then assigned to users along with a kit of tools that enable them to effectively execute the tasks assigned to them. In the case of physical products, the designs that users create using a toolkit are

> “Users’ ability to innovate is improving *radically* and *rapidly* as a result of the steadily improving quality of computer software and hardware, improved access to easy-to-use tools and components for innovation, and access to a steadily-richer innovation commons”

then transferred to producers for production (von Hippel and Katz, 2002). Toolkits make innovation cheaper for users and also lead to higher customer value. Thus, Franke and Piller (2004) in a study of consumer wrist watches, found the willingness to pay for a self-designed product was 200% of the willingness to pay for the best-selling commercial product of the same technical quality. This increased willingness to pay was due both to the increased value provided by the self-developed product and the value of the toolkit process for consumers engaging in it. (Schreier and Franke, 2004).

Producers that offer toolkits to their customers can attract innovating users into a relationship with their firms and so obtain an advantage with respect to producing what the users develop. The custom semiconductor industry was an early adopter of toolkits. In 2003, more than $15 billion worth of semiconductors were produced that had been designed using this approach (Thomke and von Hippel, 2002).

Innovations developed by users sometimes achieve widespread dissemination when those users become producers—setting up a firm to produce their innovative product(s) for sale. Shah (2000) showed this pattern in sporting goods fields. In the medical field, Lettl and Gemunden (2005) have shown a pattern in which innovating users take on many of the entrepreneurial functions needed to commercialize the new medical products they have developed, but do not themselves abandon their user roles. New work in this field is exploring the conditions under which users will become entrepreneurs rather than transfer their innovations to established firms (Hienerth, 2004; Shah and Tripsas, 2004).

I summarize this overview article by saying again that users’ ability to innovate is improving *radically* and *rapidly* as a result of the steadily improving quality of computer software and hardware, improved access to easy-to-use tools and components for innovation, and access to a steadily-richer innovation commons. Today, user firms and even individual hobbyists have access to sophisticated programming tools for software and sophisticated CAD design tools for hardware and electronics. These information-based tools can be run on a personal computer, and they are rapidly coming down in price. As a consequence, innovation by users will continue to grow even if the degree of heterogeneity of need and willingness to invest in obtaining a precisely-right product remains constant.

Equivalents of the innovation resources described above have long been available to a few within corporations. Senior designers at firms have long been supplied with engineers and designers under their direct control, and with the resources needed to quickly construct and test prototype designs. The same is true in other fields, including automotive design and clothing design: just think of the staffs of engineers and model makers supplied so that top auto designers can quickly realize and test their designs.

But if, as we have seen, the information needed to innovate in important ways is widely distributed, the traditional pattern of concentrating innovation-support resources on a few individuals is hugely inefficient. High-cost resources for innovation support cannot efficiently be allocated to "the right people with the right information": it is very difficult to know who these people may be before they develop an innovation that turns out to have general value. When the cost of high-quality resources for design and prototyping becomes very low (the trend we have described), these resources can be diffused very widely, and the allocation problem diminishes in significance. The net result is a pattern in which development of product and service innovations is increasingly shifting to users—a pattern that will involve significant changes for both users and producers.

REFERENCES

Allen, R. C. (1983), "Collective Invention", *Journal of Economic Behavior and Organization* 4(1), pp. 1-24.

Arrow, Kenneth J. (1962), "Economic Welfare and the Allocation of Resources of Invention", *The Rate and Direction of Inventive Activity: Economic and Social Factors*, A Report of the National Bureau of Economic Research, 609-625, Princeton, NJ: Princeton University Press.

Arundel, A., and V. Sonntag (1999), "Patterns of Advanced Manufacturing Technology (AMT) Use in Canadian Manufacturing: 1998 AMT Survey Results", Research Paper no. 12, Ottawa: Science, Innovation and Electronic Information Division, Statistics Canada.

Benkler, Y. (2002), "Intellectual Property and the Organization of Information Production", *International Review of Law and Economics* 22(1), pp. 81-107.

Bessen, J. (2003), "Patent Thickets: Strategic Patenting of Complex Technologies", Research on Innovation, Boston University School of Law, Working Paper.

de Jong, Jeroen P. J., and Eric von Hippel (2009), "Measuring User Innovation in Dutch High Tech SMEs: Frequency, Nature and Transfer to Producers", MIT Sloan School of Management Research Paper No. 4724-09 (March), *Research Policy,* forthcoming http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1352496

Foray, D. (2004), *Economics of Knowledge,* Cambridge, MA: The MIT Press.

Franke, N., and F. Piller (2004), "Value Creation by Toolkits for User Innovation and Design: The Case of the Watch Market, *Journal of Product Innovation Management* 21(6), pp. 401-415.

Franke, N. and H. Reisinger (2003), "Remaining Within Cluster Variance: A Meta Analysis of the 'Dark' Side of Cluster Analysis", *Vienna Business University,* Working Paper.

Franke, N., and S. Shah (2003), "How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users", *Research Policy* 32(1), pp. 157-178.

Franke, N., and E. von Hippel (2003a), "Finding Commercially-Attractive User Innovations", MIT Sloan School of Management Working Paper No. 4402-03.

Franke, N., and E. von Hippel (2003b), "Satisfying Heterogeneous User Needs via Innovation Toolkits: The Case of Apache Security Software", *Research Policy* 32(7), pp. 1199-1215.

Gault, Fred, and Eric von Hippel (2009), "The Prevalence of User Innovation and Free Innovation Transfers: Implications for Statistical Indicators and Innovation Policy", MIT Sloan School of Management Working Paper #4722-09 (January).
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1337232

Harhoff, D., J. Henkel and E. von Hippel (2003), "Profiting from Voluntary Information Spillovers: How Users Benefit by Freely Revealing their Innovations", *Research Policy* 32(10), pp. 1753-1769.

Henkel, J. (2003), "Software Development in Embedded Linux – Informal Collaboration of Competing Firms", in W. Uhr, W. Esswein and E. Schoop (eds.), *Proceedings der 6. Internationalen Tagung Wirtschaftsinformatik 2003* 2, Heidelberg, pp. 1-99.

Henkel, J. and E. von Hippel (2005), "Welfare Implications of User Innovation", *Journal of Technology Transfer*, forthcoming.

Herstatt, C. and E. von Hippel (1992), "From Experience: Developing New Product Concepts via the Lead User Method: A Case Study in a "Low Tech" Field", *Journal of Product Innovation Management* 9(3), pp. 213-222.

Hertel, G., S. Niedner and S. Herrmann (2003), "Motivation of Software Developers in Open Source Projects: an Internet-Based Survey of Contributors to the Linux Kernel", *Research Policy* 32(7), pp. 1159-1177.

Hienerth, Christoph (2004) "The commercialization of user innovations: Sixteen cases in an extreme sporting industry", Vienna University of Economics and Business Administration Working Paper.

Jensen, M. C., and W. H. Meckling (1976), "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure", *Journal of Financial Economics* 3(4), pp.305-360.

Jeppesen, L. B. (2004), "Profiting from Innovative User Communities: How Firms Organize the Production of User Modifications in the Computer Games Industry", *Department of Industrial Economics and Strategy, Copenhagen Business School, Copenhagen, Denmark,* Working paper WP-04.

Lakhani, K. R. and B. Wolf (2005), "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects", J. Feller, B. Fitzgerald, S. Hissam, and K. R. Lakhani (eds.), *Perspectives on Free and Open Source Software,* Cambridge, MA: The MIT Press.

Lessig, L. (2001), *The Future of Ideas: The Fate of the Commons in a Connected World,* New York: Random House.

Lettl, C., C. Herstatt and H. Gemünden (2005), "The entrepreneurial role of innovative users", *Journal of Business and Industrial Marketing* 20(7), pp. 339-346.

Lilien, Gary L., Pamela D. Morrison, Kathleen Searls, Mary Sonnack, Eric von Hippel (2002), "Performance Assessment of the Lead User Idea Generation Process", *Management Science* 48(8) (August), pp. 1042-1059.

Lim, K. (2000), "The Many Faces of Absorptive Capacity: Spillovers of Copper Interconnect Technology for Semiconductor Chips", MIT Sloan School of Management, Working paper # 4110.

Lüthje, C. (2003), "Customers as Co-Inventors: An Empirical Analysis of the Antecedents of Customer-Driven Innovations in the Field of Medical Equipment", *Proceedings from the 32th EMAC Conference* 2003, Glasgow.

Lüthje, C. (2004), "Characteristics of Innovating Users in a Consumer Goods Field: An Empirical Study of Sport-Related Product Consumers", *Technovation*, forthcoming.

Lüthje, C., C. Herstatt and E. von Hippel (2002), "The Dominant Role of Local Information in User Innovation: The Case of Mountain Biking", MIT Sloan School, Working Paper #4377-02.

Morrison, P. D., J. H. Roberts and D. F. Midgley (2004), "The Nature of Lead Users and Measurement of Leading Edge Status", *Research Policy* 33(2), pp. 351-362.

Morrison, P. D., J. H. Roberts and E. von Hippel (2000), Determinants of User Innovation and Innovation Sharing in a Local Market", *Management Science* 46(12), pp. 1513-1527.

Nuvolari, A. (2004), "Collective Invention during the British Industrial Revolution: The Case of the Cornish Pumping Engine", *Cambridge Journal of Economics* 28(3), pp. 347-363.

Ogawa, S. (1998), "Does Sticky Information Affect the Locus of Innovation? Evidence from the Japanese Convenience-Store Industry", *Research Policy* 26(7-8), pp. 777-790.

Olson, Erik L., and Geir Bakke (2001), "Implementing the Lead User Method in a High Technology Firm: A Longitudinal Study of Intentions versus Actions", *Journal of Product Innovation Management* 18(2), (November), pp. 388-395.

Pavitt, K. (1984), "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory", *Research Policy* 13(6), pp. 343-373.

Raymond, E. (1999), *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary",* Sebastopol, CA: O'Reilly.

Riggs, William, and Eric von Hippel (1994), "The Impact of Scientific and Commercial Values on the Sources of Scientific Instrument Innovation", *Research Policy* 23 (July), pp. 459-469.

Rosenberg, N. (1976), *Perspectives on Technology,* New York: Cambridge University Press.

Rosenberg, N. (1982), *Inside the Black Box: Technology and Economics,* New York: Cambridge University Press.

Schreier, M. and N. Franke (2004), "Tom Sawyer's Great Law in Action: Why Users are Willing to Pay to Design their own Products via Toolkits for User Innovation and Design", Working Paper, Vienna University of Economics and Business Administration.

Schumpeter, Joseph A., *The Theory of Economic Development,* New York: Oxford University Press, 1934.

Shah, S. (2000), "Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment", MIT Sloan School of Management, Working paper # 4105.

Shah, S., and M. Tripsas (2004), *When Do User-Innovators Start Firms? Towards A Theory of User Entrepreneurship,* University of Illinois, Working Paper #04-0106.

Shapiro, C. (2001), "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting", A. Jaffe, J. Lerner, and S. Stern (eds.), *Innovation Policy and the Economy* 1, pp. 119-150, Cambridge, MA: The MIT Press.

Stigler, George J. (1951), "The Division of Labor is Determined by the Extent of the Market", *Journal of Political Economy* 59(3) (June), pp. 185-193.

Thomke, S. H., and E. von Hippel (2002), "Customers as Innovators: A New Way to Create Value", *Harvard Business Review* 80(4), pp. 74-81.

Urban, G. L. and E. von Hippel (1988), "Lead User Analyses for the Development of New Industrial Products", *Management Science* 34(5), pp. 569-82.

Varian, H. R. (2002), "New Chips Can Keep a Tight Rein on Consumers", *New York Times,* July 4.

von Hippel, E., and S. N. Finkelstein (1979), "Analysis of Innovation in Automated Clinical Chemistry Analyzers", *Science & Public Policy* 6(1), pp. 24-37.

von Hippel, E. (1994), "Sticky Information and the Locus of Problem Solving: Implications for Innovation", *Management Science* 40(4), pp. 429-439.

von Hippel, E., and G. von Krogh (2003, "Open Source Software and the 'Private-Collective' Innovation Model: Issues for Organization Science", *Organization Science* 14(2), pp. 209–223.

von Hippel, Eric (2005), *Democratizing Innovation*, Cambridge MA: The MIT Press.

von Hippel, E., and R. Katz (2002), "Shifting Innovation to Users via Toolkits", *Management Science* 48(7), pp. 821-833.

BBVA

The Power of Creative Freedom: Lessons from the MIT Media Lab

Frank Moss
MIT Media Lab

Over its 25-year history, the MIT Media Lab has refined a unique research style that has resulted in some of the most original thinking of the digital revolution. The secret formula for this success–and for the Lab's continued ability to "invent the future"–is a renegade research environment that not only allows, but encourages, researchers to ask the questions that no one else has thought to ask.

We are an organization that is constantly reinventing itself with the most unconventional pairings of disciplines and people. The idea is to bring together the brightest people we can find–from a large number of disparate disciplines–to figure out to how to change the world. Most important, we are simply not afraid of being on the "lunatic fringe" or hitting a dead end. That's because real innovation comes from fostering a research culture where it is not only okay to fail, but where failure is fully expected–and accepted–as part of the creative process. Great ideas don't come from playing safe. They don't come from thinking incrementally. Rather, they come from thinking about things in a way that no else has. And this is the lesson that the Lab offers to the world.

I believe that all organizations, both for-profit and non-profit, can learn from the Lab's unique ethos as long they are willing to take the risk of doing things a bit differently.

SEEDING INVENTION

Drawing on the Media Lab's example, if you were to ask me where you might strengthen an organization's innovation process, I'd say to begin by rethinking how you define innovation. Too often, the early seeds of creativity are undervalued. It is commonly accepted that innovation is the successful *implementation* of creative ideas. But, as we at the Media Lab have demonstrated, true innovation isn't about finding new ways to put existing ideas into new practices. The process needs to start much earlier, and be far more radical. It needs to begin as "pre-innovation"–with crazy, revolutionary ideas that become the fodder for society-changing technologies and products.

Today, too many companies are weak on the front end of the innovation cycle because they are not investing in the seed corn—those hundreds of inventions that result from a free-formed, undirected process.

Think of the kind of innovation that, in the past, came from Bell Labs, IBM, or Xerox PARC. These companies made a conscious decision to invest heavily in seeding new ideas; some would lead to products, but many more would not. However, the entire organization was dedicated to pioneering new ways of thinking about technology. Now, we're seeing more and more organizations delegating "innovation" to smaller, elite teams of creative thinkers who reside in a company's "innovation lab." This is a model that does not take advantage of the power of creative freedom. Innovation should be ubiquitous throughout an organization. Ideas should organically sprout from all different departments in such a way that there is no "wrong way" to think about a problem, or "right way" to solve one.

FOLLOW YOUR PASSION

At the Media Lab, our mantra is "follow your passion." We're not here to answer specific questions for our sponsors or outside funding agencies, but rather to discover the new questions that need to be asked–to focus on how digital technology can help to transform our most basic notions of human capabilities. Most important, we are here to foster a unique culture of *learning by doing.* To do this, we have gathered some 25 research groups working "atelier style" to create the things that conventional wisdom says can't–or shouldn't–be done. There are no boundaries, only possibilities.

REACHING ACROSS TRADITIONAL DISCIPLINES

Central to the Media Lab culture is our disregard for working within the straightjacket of traditional academic disciplines. Here, for example, the Opera of the Future research group, which is expanding the boundaries of music and creativity, shares a lab with the Smart Cities group, which is focused on designing tomorrow's sustainable cities. The Tangible Media group, which focuses on tactile connections between the physical and digital worlds, works alongside researchers in Viral Communications who are exploring radical new concepts for networked systems.

Research disciplines at the Lab range from robotics, to neurobiology, to epistemology. And it is not unusual for any single research project to draw from ongoing work in several seemingly unrelated disciplines: the challenge is to find the connections. Each research group is led by a faculty member or research scientist who directs a team of graduate and undergraduate students. (The undergraduates work at the Lab through MIT's Undergraduate Research Opportunities Program.) All researchers also work within one of the Lab's consortia, which are organized around broad research themes rather than traditional disciplines. For example, Things That Think, the Lab's largest consortium, joins computer scientists with product designers, biomedical engineers, and even architects to focus on inventing the future of digitally augmented objects and environments.

The Lab's 25 research groups focus on more specific areas, and address a broad scope of projects that range from creating the next generation barcode (Camera Culture), to developing communications systems with an understanding of their content (Object-Based Media), to creating interfaces that allow people to "grasp" and "manipulate" bits by coupling them with everyday objects and architectural surfaces (Tangible Media).

One key feature of this approach involves shunning the standard academic model of directed research. The Lab's funding model gives all corporate sponsors access to *all* of the Lab's intellectual property during their term of sponsorship, license-fee free and royalty free. This further promotes the intellectual openness and sharing that is essential to the way the Lab functions. Each faculty member or student has total freedom to stray from conventional research paths and collaborate with others in entirely different areas. This way, the Lab becomes an open-ended think tank, with access to cutting-edge research on a wide variety of topics in many different fields.

SERENDIPITY BY DESIGN

When you create the right research culture, some of the best innovation happens through the most serendipitous paths–"accidentally on purpose." For example: in the 1990s, the Lab's Physics and Media group began exploring the interaction between the human body and electrical fields while developing new sensors for a collaboration with world-renowned cellist Yo-Yo Ma. This led to smart furniture that could "see" in 3D, and to the subsequent discovery of a way to send data through the human body, which was incorporated into a Spirit Chair for magicians Penn and Teller (who were appointed as visiting scholars at the Lab). The device literally channeled a field through a performer's body to control music.

One day, while watching Penn and Teller demonstrate the Spirit Chair, a visiting executive from Lab sponsor NEC got the idea that this same technology could be used for a car-seat sensing device. Once its potential use was identified, the Media Lab was instrumental in quickly demonstrating the product's technical viability, and in helping to understand the physics of the problem, as well as in helping with the project's engineering design and rapid prototyping. The executive had a prototype car-seat detector built at the Lab and demonstrated it to customers within the same year. Soon afterward NEC announced the Passenger Sensing System. Armed with this technology, the car's seat could distinguish between a rear-facing or forward-facing baby, and could signal an auto's airbag when–and more importantly when *not*–to deploy, making it a potentially life-saving device for a baby traveling in the front seat of an automobile.

Another example of how an open, creative environment can bring about surprising results involves the Lab's work in affective computing. The initial focus of this work was on developing computers with "emotional intelligence" to detect user frustration in human-machine interaction. Over time, it has grown into two very important, yet totally different research areas: devices used to detect and respond to a customer's level of satisfaction, and tools to help detect and treat autism.

The potential for using affective computing for customer service is significant. Currently, there is no existing system to capture and analyze facial expressions in real-time customer-service interactions and relate these to business outcomes. But consider how important it is to a customer-service oriented business, such as banking, to have real-time techniques to assess the outward appearance of customer interest, confusion, and other cognitive-affective states. This capability could lead to fundamental new understanding of how to improve customer experience in face-to-face interactions, at ATMs, or through online banking services. This technology can also be used for more accurate results from test marketing, where participants are often less than honest when filling out surveys or being interviewed. It can also help take voice emotion out of consumer phone interactions, helping to diffuse difficult interactions with a customer-service representative.

But this same technology also promises to have an impact in helping people with autism. Specialized tools, including novel, wearable physiological sensors and corresponding software, can help individuals on the autism spectrum communicate cognitive and emotional states, as well as help others–including scientists, therapists, teachers, and caregivers–to better understand those states.

RETHINKING TRADITIONAL RESEARCH BOUNDARIES

Linus Pauling once said, "The best way to have a good idea is to have a lot of ideas." To my mind, no one demonstrated this more than the late William J. Mitchell, former dean of MIT's School of Architecture + Planning and head of the Media Lab's Smart Cities research group. Bill, who sadly passed away a few months ago, was an inspirational thinker and prolific writer who challenged conventional concepts about sustainable cities, design, and urban transportation by bringing together the most unlikely team of researchers. His Smart Cities research group continues to design the CityCar, a light-weight, electric, shared vehicle that folds and stacks like a supermarket shopping cart and would be placed at convenient locations throughout urban areas. The CityCar project totally rethinks the design of an automobile, as it has all essential mechanical systems housed in the car's wheels. The car itself is amazing. But even more amazing is the fact that the CityCar team does not include one researcher who had a background in automobile design.

In addition to crossing traditional disciplinary lines, innovation also involves breaking out of conventional thinking about what kind of researchers should be approaching a particular problem. What would have happened if we had depended on typewriter companies to come up with word processing? Or if we didn't look beyond landlines for the next breakthrough in telephony?

Health care offers another good example. At the Media Lab, engineers, scientists, and designers, unencumbered by current industry and academic biases, are exploring a myriad of health-related issues. With "out-of-the-box" thinking, Lab researchers have already made major strides in developing new "smart" prostheses for amputees, memory aids, and even an ingenious new technology for analyzing and precisely controlling any neural circuit, including those in the brain. This new work in neuroengineering offers the possibility of controlling the firing of specific neurons in the brain to within a millisecond, very precisely targeting cells so that neighboring healthy cells remain untouched. This work has implications for developing radically new medical technologies to treat brain disorders such as Parkinson's disease, or even blindness, and for changing mental and emotional states, such as severe depression.

Lab researchers are also focusing on a new area we call New Media Medicine, where we seek to shift the health-care paradigm. We believe that to have a truly meaningful impact on society, health must be approached in a far broader context—a context in which an individual's physical, mental, and social well-being are so closely integrated that they cannot be distinguished. Toward this end, the Media Lab is designing new platforms and applications that will become intimate yet unobtrusive parts of a person's everyday life. These range from next-generation smart phones and personal sensing networks to help their users become more self-aware and understand how to adapt better behaviors

“The Media Lab encourages all its researchers never to abandon their child-like fascination with the universe or their thrill of discovery. Not only do we stress the importance of learning by doing, but also the importance of having fun during the process”

for healthier lifestyles, to technologies for “personal collective intelligence,” where people can contribute to, and learn from, the collective knowledge and experience of their peers. We are also developing digital tools for helping patients become equal partners with their physicians–allowing them to share and interpret their health information to positively change their lives for the better.

HARD FUN

Many years ago, the Lab adopted its now iconic motto of “demo or die.” Another expression we use to describe our unique culture is “imagine and realize.” Our students are constantly challenged to build, and build again, and then to demonstrate their work at scale. We are a Lab of tinkerers. It would not be unusual for a visitor to see a sewing machine or a soldering iron sitting next to a state-of-the-art digital display. One day a student is busily cutting out a cardboard model; the next day that same student is writing complex code.

The Media Lab encourages all its researchers never to abandon their child-like fascination with the universe or their thrill of discovery. Not only do we stress the importance of learning by doing, but also the importance of having fun during the process.

In fact, one of the Media Lab’s research groups is called Lifelong Kindergarten. A great name, but an even greater approach. Here, researchers develop new technologies that, in the spirit of the blocks and finger-paint of kindergarten, expand the range of what people design, create, and invent–and what they learn in the process. Their ultimate goal is a world full of playfully creative people who are constantly inventing new possibilities for themselves and their communities. One recent innovation that has come out of the group is Scratch, a programming language and online community that makes it easy to create interactive stories, games, animations, and simulations–and share these creations online. Scratch is designed to enhance the technological fluency of young people (ages eight and up), helping them learn to express themselves creatively with new technologies. As they create and share Scratch projects, young people learn to think creatively, reason systematically, and work collaboratively. Available at no cost through the Internet, Scratch has reached a broad, worldwide audience with more than 500,000 users who have uploaded more than one million projects.

A UNIQUE PHYSICAL ENVIRONMENT

In March 2010, the Media Lab expanded into a spectacular new building designed by Fumihiko Maki, who like I. M. Pei (designer of the original 1985 Media Lab building), is a Pritzker Prize-winning architect. This new space is a model for how physical space can be fully integrated with a research program and serve not just as a place to work, but also as a catalyst for creativity. The entire complex functions as an evolving research platform, seamlessly connecting the real and virtual worlds.

The building offers new levels of transparency, where all the researchers can view each other from various vantage points, supporting the unfettered exchange of ideas. The goal is a space that functions as a single, massively interconnected unit, with seven labs facing one another across a central atrium in a staggered configuration. Through a series of interactive information displays placed throughout the Media Lab complex, the Lab expands collaboration beyond our walls to visitors, sponsors, colleagues, and the public at large. It delivers on our unique vision of how to conduct society-changing research—no boundaries, no walls—just a flow of interdisciplinary ideas, and plenty of open space to invent just about anything.

SYMBIOSIS

The Media Lab model offers enormous freedom to pursue the most far-out research directions without concern for the accepted conventional approach in either academia or industry. At the same time, the Lab's close connections to the corporate community (both through sponsor visits to the Lab and faculty and student visits to sponsors' research facilities), help keep research grounded in real-world concerns.

The Lab and its sponsors have a symbiotic relationship: the Lab's open research approach enables companies to ask questions they would not have otherwise asked. The goal is for collaboration with the Media Lab to help widen the front end of a company's R&D pipeline and spur innovative thinking about entirely new directions. At the same time, the Lab's interaction with industry helps it stay connected with real-world needs.

It is important to note that the Lab does *not* focus on specific, technology-based projects for our sponsors, but rather looks to create an environment for companies to improve *their own* innovation processes. If a sponsor is using the Lab correctly, it will come in looking for a single solution and walk away with ideas that relate to five entirely different areas of product development. The goal is to fuel the imagination–to encourage thinking outside the box; for companies to become visionary about their research direction.

This is not just academic theory; the Media Lab has lived this for 25 years. Today the same Lab that predicted the convergence of multimedia and technology, and paved the way for the digital revolution in 1985, continues to break new ground with society-changing advances. More than 80 spin-off companies have come out of the Lab, and Lab-based commercial products range from electronic ink (the basis of the Kindle), to LEGO Mindstorms, to Guitar Hero. We have some 60 sponsors that include a number of the largest and most prestigious companies in the world, including Audi-Volkswagen, AOL, BT, BBVA, Bank of America, Google, IBM, Intel, LEGO, Samsung, Sun Microsystems, and Toshiba.

MAKING A DIFFERENCE

For years, technologists digitized almost everything, but *transformed* almost nothing. Now we are moving away from merely building more sophisticated digital tools, and are looking to create technologies that will be truly intelligent and helpful participants in the world.

Like the Media Lab's past work, today's research remains clearly focused on human experience. But more than ever before, it emphasizes the strong link between business, society, and the individual. A few examples of some current projects that are "making a difference" include:

- HealthMap's Outbreaks Near Me iPhone and Android application, which provides real-time disease outbreak information.
- Mobility on Demand (MoD) systems–lightweight, electric vehicles placed at electrical charging stations strategically distributed throughout a city. MoD systems provide mobility from transit stations to and from a final destination. Three MoD vehicles have been developed: the CityCar, RoboScooter, and GreenWheel bicycle.
- CollaboRhythm, a speech- and touch-controlled collaborative interface that facilitates improved doctor-patient interaction. Patients can actively engage with their data, allowing for a more collaborative relationship with their doctors.
- Konbit, a mobile phone-based system that helps communities rebuild themselves by soliciting skill sets of local residents. The system, which does not require participants to be literate, indexes the skills of all those who phone in, translates the responses to English, and makes them searchable by NGOs via natural language processing and visualization techniques.
- Sourcemap, a volunteer-driven, social-networking Web application that presents easy-to-understand map visualizations of the environmental impact of consumer products—information that is almost never available to the public.
- NETRA (Near-Eye Tool for Refractive Assessment), a quick, simple, and inexpensive way for people in the developing world to use mobile phones to give themselves eye exams. A small plastic device–which currently can be produced for less than US$2–is easily clipped onto a mobile phone screen. To use it, one simply holds the device up to the eye, looks into it, and uses the phone's keypad until two patterns overlap. This is repeated several times per eye, with the patterns at different angles. The whole process takes about two minutes, during which time software loaded onto the phone provides the data needed to create a prescription.

The Lab offers us all an outstanding model for how much an organization can accomplish when it fosters an environment where people can create at will, follow their passion, and think beyond the boundaries set by conventional thinking. The sky is the limit when no one tells you that "it can't be done." We *can* invent our own future.

Noted Accomplishments from the MIT Media Lab

E-ink, opening up the possibility of a one-book library.

SixthSense, a gestural, pendant-like interface that projects digital information onto any surface, and allows the user to interact with that information using natural hand gestures. It seamlessly integrates information with the user's physical surroundings, making the entire world a computer.

Scratch, an open-source programming language for kids that allows them to create their own interactive stories, games, music, and animations for the Web.

CityCar, a shared, foldable, electric, two-passenger vehicle for urban areas.

The world's first powered ankle-foot prosthesis, an important advance for lower-limb amputees. The device propels users forward using tendon-like springs and an electric motor, reducing fatigue, improving balance, and providing a more fluid gait.

Nexi, a social robot that possesses a novel combination of mobility, moderate dexterity, and human-centric communication and interaction abilities.

The world's first real-time, moving synthetic hologram.

Bokode, a next-generation barcode that uses a new optical solution for encoding information, allowing barcodes to be shrunk to fewer than 3mm, read by ordinary cameras, and offer different information from different angles.

Csound, a pioneering computer programming language for transmitting music over the web. It is one of the most widely used software sound systems.

The first Web-based, on-demand, personalized electronic newspaper.

The first "programmable brick" that led to the LEGO Mindstorms robotics kits, now used by millions of people around the world.

The XO machine, known worldwide as the "$100 Laptop," which offers connectivity to children throughout the developing world.

Sensors that can detect a user's actions by measuring a body's influence on an electric field.

Audio Spotlight, that generates audible sound that can be directed to one specific location.

The MIT Media Lab's Innovative Evolution

In 1985, MIT Professor Nicholas Negroponte and former MIT President Jerome Wiesner co-founded the Media Lab, which grew out of work of MIT's Architecture Machine Group. Cross-disciplinary in nature, the Lab housed researchers in fields ranging from holography, to documentary film making, to epistemology and learning. Not only encouraged to think "at the lunatic fringe," Lab researchers were also encouraged to build prototypes of their ideas. Rather than the standard academic theme of "publish or perish", the Lab's motto was "demo or die".

The physical environment of the Lab supported this unconventional thinking. Housed in a building designed by Pritzker Prize-winning architect I.M. Pei, the Lab pioneered the concept of open computer gardens, with personal computing on every desk and an eclectic range of ongoing projects on display. Glass-fronted offices ringed the perimeter of each floor. One lab would be full of LEGOs, while another contained the most sophisticated equipment for holography. Lab visitors came not only to see what Lab researchers did, but also *how* they did it.

The Lab's initial research focus was often represented as a Venn diagram representing publishing, cinema, and computers. From its first days, there was a focus on computing for people during a time when no one was thinking in terms of "user-friendly" machines or why our machines needed to be adapted to human ways.

During its second decade, the Lab's research foci morphed into a revised Venn diagram showing the convergence of perceptual computing, learning and common sense, and information and entertainment. The overlapping segments were labeled interact, play, and express, and the Media Lab was shown in the center of this convergence.

For the first time, the idea of user-friendly was greatly expanded. The Lab extended its interests to focus more broadly on pervasive, ubiquitous computing. The Lab began looking at merging the virtual and physical worlds—initiating work to integrate emerging digital technologies into everyday objects. The Lab began to present such seemingly outlandish concepts as a refrigerator that could tell you when you were low on milk, or a car that could give you directions, or point out a good restaurant that you'd pass on your way home. During this time, the Lab also pioneered wearable computing–the idea that we could actually wear our bits on our clothing or carry them in our bags. It also conducted groundbreaking research in sociable and tangible media, further enhancing individual and community expression and social connection.

Now in our third decade, the Lab has added human augmentation as a major research theme. At some point in our lives, almost all of us will be marked by a fundamental disability, from dementia, to the loss of a limb, to a debilitating disease such as Parkinson's. Indeed, serious physical and mental challenges are inherent to the human condition. But the Lab does not believe that we need to accept the current definition of disability. Instead we are asking, "What if, through the invention of novel technologies, we could profoundly improve the quality of life for those afflicted with physical, cognitive, or emotional disabilities, while significantly reducing health-care costs? What if natural ability was a baseline, and enhanced ability became the norm?"

BBVA

Creating Abundance through the application of a discipline of innovation

Curtis R. Carlson
SRI International

I. THE INNOVATION ECONOMY

Innovation is the creation and delivery of new customer value into the marketplace.[1] It is the only path to growth, prosperity, environmental sustainability, and security (Carlson and Wilmot, 2006). Developed countries can no longer compete on the basis of low-cost labor or access to capital, which flows freely around the globe. They must provide an environment that promotes continuous and efficient innovation. This is the only way for developed countries to remain productive and competitive, with increasing personal incomes and high levels of employment.

Today, many companies are doing poorly at innovation. To thrive, companies need new innovation perspectives and skills. They must embrace a broader, more comprehensive understanding of their opportunities for creating customer value. This broader understanding emphasizes the importance of continuous value creation throughout all parts of the enterprise to remain competitive. With such skills, the future can be seen correctly as a period of abundance (Carlson and Wilmot, 2006: 22). Without them, the future may be seen correctly as a period of scarcity.

Of course, innovation has always been the driving force for progress and improved productivity (Ridley, 2010). What is different today is the intensity of the innovative processes needed to sustain enterprises and national competitiveness. Specifically, the innovation economy is characterized by three main attributes (Carlson and Wilmot, 2006: 26; parts of this article were abstracted from Carlson and Schaufeld, 2011).

Abundance of Opportunities: This is a time of unprecedented opportunity. Almost every major field is undergoing increasingly rapid technological development. Progress is often at exponential rates, with improvements of 100% at the same cost every 12 to 48 months (Kurzweil, 2005). The Moore-Engelbart Law[2] for computers is the most famous example of this property. However, rapid, exponential improvement is now seen in many other fields too, as they become increasingly based on ideas and bits, not just atoms and muscle. New ideas are the currency of the innovation economy, and they are an abundant, unlimited resource.

These continuous, rapid improvements open up one major opportunity after another. Whether in finance, medicine, media,

[1] A more comprehensive and inclusive definition is: "Innovation is the creation and delivery of new customer value in the marketplace. Innovations are sustainable only if they provide sufficient enterprise value to allow for their continued production."

[2] J. Markoff (2005) tells how Moore heard a talk by Douglas Englebart about why, because of basic scaling principles, computers would improve at these rates. Moore then plotted the data and created the concept that now carries his name.

energy, consumer electronics, computing, or communications, there has never been a better time for creating major new innovations. It is potentially a time of great prosperity—but only if we seize and address the innovation challenge.

Consider, for example, access to financial services. For most consumers, the knowledge needed to understand and access the wide range of options available is daunting. But increasingly there will be computer "assistants" to help customers with these options. There are primitive versions of these computer assistants available now on smart phones. But they will quickly become impressively more "intelligent" and allow for a host of convenient, instantaneous banking transactions.

Creation and Destruction of Companies: While rapid, exponential progress creates great opportunities, it also creates great challenges. A company that does not innovate at the speed of its market and does not adapt to technological change will decline. The decreasing life span of S&P 500 companies indicates that fewer of them are keeping pace with change (Foster and Kaplan, 2001). "Lifetime employment" has become a distant, quaint idea in many parts of the world.[3] If history is a guide, new players will arise who understand these opportunities and move rapidly to displace today's leaders. One example is what is happening to bookstore retailers. Online retailers, such as Amazon.com and Kindle-like digital readers, are replacing them. A similar fate awaits video rental companies such as Blockbuster, which is now contemplating bankruptcy, as their brick-and-mortar store advantage almost literally turns to dust before their eyes. Will this happen to conventional banking too? There are already many companies, such as PayPal, working to remove conventional banks from transactional processes.[4]

At the same time, the opportunity to create world-leading companies has never been greater. Google was started little more than 10 years ago by two students with an idea. At this point it is a $144B company that dominates its industry. AOL, Yahoo, eBay, and Amazon all had similar origins. In fact, it can reasonably be said that the "old" industries of media, banking, pharmaceuticals, education, energy, and many others, are all destined to follow the well-known path of creative destruction and then re-emerge as new, major industries.

Intense Global Competition: The world is now deeply integrated, and competition is increasing at an unprecedented rate. Almost every significant business must now think and act globally in our "flat world," where ideas and money travel at the speed of light (Friedman, 2005). Countries like India and China are rapidly moving past low-cost labor alone as a competitive advantage, because they can leverage the entire world's knowledge. They can bring proven business ideas and technologies into their countries and adapt them for regional markets. It is possible to argue that China is now the leading innovation country in the world. China is taking established businesses from the West, modifying them to fit the Chinese ecosystem and, at the same time, developing new products, services, and models of production. In 2010 China passed Japan in GDP (Hosaka, 2010) and it is now the world's largest and fastest growing automotive market.[5]

Consider also that based on its population alone, China has the potential for more "honor students" than America has students.[6] It is perhaps not surprising that China and India together annually produce more than

[3] Note: The *velocity* of technological improvement at rapid, exponential rates also implies the *acceleration* of technological improvement at rapid, exponential rates. This is a sobering realization, the consequences of which for individuals, businesses, and nations are surely impossible to fully appreciate.

[4] Wikipedia, http://en.wikipedia.org/wiki/PayPal

[5] How China will Change the Cars America Drives, Motor Trend, April 25, 2010, http://mt.kargo.com/v/News/HowChinaWillChange/?KSID=3189d3546687c862a6eebeb2eaf0ef7b

[6] See for population statistics the *CIA Fact Book* at https://www.cia.gov/library/publications/the-world-factbook/

10 times as many science and engineering graduates as the United States. Although the quality of America's graduates still puts the United States ahead, this advantage may not last long (Wadhwa, 2005). In India and China, a fervent desire for education along with prodigious work ethics and cultures of entrepreneurism create a strong basis for rapid progress.

At the same time, we should be cautious about predicting China's long-term prospects, since we have neither full access to information about their economy nor the ability to predict the future path of their political system (Friedman, 2009). India, with all its promise, must address daunting infrastructure, environmental, and governance issues (Kapor, 2010). But clearly, increased levels of global competition have emerged. Imagine what global competition would be like if the nearly four billion people now living in poverty across India, China, and the other developing countries fully join the world's economy and add their ideas, energy, and innovative genius.

Other Issues: The innovation economy has other special challenges. Environmental costs are increasing. Additionally, the cost of fighting terrorism is unabated, taking resources away from other activities. It is impossible to anticipate what future terrorist events might do to open societies, from the loss of personal freedoms to restrictions on business interactions. In 2010, the world is emerging from a period of financial chaos, but it is still not clear whether institutional changes made in response to the crisis will help or hinder future growth.[7]

Finally, there are major demographic shifts occurring around the world whose consequences are not fully understood. For example, in Germany, France, Italy, Japan, Korea, Singapore, and many other developed countries, indigenous populations are declining by 25% to 50% in each successive generation.[8] This is also true in China because of their one-child policy. In the future, without effective immigration policies, there may be many fewer workers in these countries to support the costly social services required for increasingly older populations.

For all these reasons, to thrive we must significantly improve our success rate in all forms of innovation. It is the *only* factor that significantly counteracts these rapidly increasing costs and other complex challenges. For executives in companies, the innovation economy forces management to increasingly shift its focus from gradual improvement of current assets to the creation of new, high-value products and services. The daunting rate of change of both technologies and markets demands this shift in emphasis.

II. THE OPPORTUNITY TO IMPROVE PERFORMANCE

Considerable attention is being given to the topic of innovation. A Google search query on "innovation" produces more than 100 million results. The concept has become a source of theory, research, scholarly writing, and endless discussion in the press. There is a litany of consultants, publications, and public conversation about the virtues of innovation as a strategy. In the U.S., the government has established a new National Council on Innovation and Entrepreneurship.[9]

But something is still missing. Michael Mandel, chief economist at *Bloomberg Business Week*, wonders why, with our wide array of nanotechnology, biotechnology, robotics, artificial intelligence, and other technologies, we are not seeing more marketplace impact (Mandel, 2009). He further asks why we don't have better tools for quantifying progress. We have output

[7] In addition, we are rapidly going from 1.5 billion Internet users today to a time, only years from now, when a large percentage of the world's population of seven billion people will be connected. Individual connections are only one important development, however. Additional computer applications will be connected via the Internet and run at many millions of times today's computer speeds. Systems of all types — financial trading, consumer services, production design systems, etc. — will be orders of magnitude more plentiful and complicated when compared to those available today. No person or enterprise will be capable of understanding all of them. Indeed, the behavior of these systems will be non-linear, and they will interact in ways that can neither be tested nor anticipated. Given this complexity, the large number of computer hackers, and the criminals supported by nation states working to destroy or extract value, we should expect that "Black Swans" will be even more common (Taleb, 2007).

[8] "Population Decline," *Wikipedia*, http://en.wikipedia.org/wiki/Population_decline

[9] The author of this article, Curtis R. Carlson, is a member of this council. See http://www.commerce.gov/news/press-releases/2010/07/13/locke-announces-national-advisory-council-innovation-and-entrepreneur

measures, such as the number of initial public offerings (IPOs), stock price, corporate growth, and market share. He argues that these measures fall short because they do not measure either innovative capacity or efficiency. Measuring the number of patents or publications has not proven to be particularly effective.

Marketplace output is the only true measure of innovative effectiveness. However, innovative progress, capacity, and efficiency can be measured using the "artifacts" of innovation, such as the core concepts and processes to be described shortly. Innovation will be faster and more successful once these core concepts are widely understood and applied.

Poor Innovative Performance: The lifetimes of the largest companies in America are decreasing rapidly. At the turn of the 20th century, a large company would continue to be included in the S&P 500 index of large-cap American stocks for more than 75 years before it was bought or went away. Today, the lifespan of this elite group of companies is, on the average, down to less than 20 years (Foster and Kaplan, 2001; Carlson and Wilmot, 2006: 34). These companies, with all their advantages, are not keeping pace. They are like dinosaurs whose bulk, once an advantage, has become a disadvantage since it fatally slows down their ability to adapt. Today, it takes different processes and corporate architectures to survive.

Consider also the success rate of new products in the retail grocery industry, which is only 20 to 30% (Stone, 2008). Do they fail because of bad technology or from lack of clever ideas? No. They fail because customers do not want them. Even in Silicon Valley, by far the world's leading new venture creation region, only one out of seven or ten new companies has real success. In what other activity would this be seen as good performance?

Example after example can be given for innovative failure. This quote is indicative of the problem: "If you ask a CEO whether the world is moving faster and whether they need to innovate faster, they will say yes. But if you ask an employee in that company to describe their innovation system, you get blank looks. They have none" (Carlson and Wilmot, 2006). My organization, SRI International, has worked with hundreds of companies and organizations, and that is also our observation. Most organizations do not have comprehensive innovation systems or processes. If the professionals in a company cannot describe the company's innovation processes, there clearly are none.

University technology transfer programs are also often considered to be disappointing in generating value from their intellectual property (Mitchell). Much of it lies fallow. Universities are, of course, not designed to create innovations. Their mission is education and the generation of new knowledge. Nevertheless, universities have built into their technology transfer programs an important flaw. If there is one thing we know about innovation, it is that "technology push" does not work. Rather, the goal should always be "market pull." University technology transfer initiatives are mostly technology push. If these programs are to be improved, they must reverse this approach and create incubators that focus on "value creation"—that is, formally and systematically connecting market needs with new solutions.

The amount of waste these failures represent is enormous. Today's low output of innovations is analogous to the low quality and high cost of products in the 1950s. Imagine if we performed just a few percent better every year in our innovative ability.

Over time, the positive impact of these improvements on companies and national economies would be enormous.

III. VALUE—NOT ONLY COST AND QUALITY

Given these dynamics of the innovation economy, are companies and their workforces fully preparing to compete? Enterprises that do not strengthen and broaden their innovation processes will fail. On the other hand, individuals who master such skills will be uniquely valuable. To gain a perspective on the potential of the innovation economy for improvement, it is useful to look at an example from a previous economic period that illustrates the enormous improvements possible when people work in more productive ways.

In the 1960s and 1970s, America lost its lead as producer of quality products to Japan. After World War II, a "Made in Japan" label implied cheaply made goods. Japanese companies were determined to eliminate that perception. They accomplished this by embracing the Total Quality Management (TQM) movement, as pioneered by W. Edwards Deming (1986) and Toyota's Taiichi Ohno (1988). These innovators proved that by working in a new, more productive way based on fundamental improvement principles, companies could dramatically increase quality *and* dramatically reduce costs. Using Ohno's lean manufacturing innovations, Toyota became the world's leader in automotive quality and eventually the world's number one car company. [10]

At first, the US and other developed countries ignored Japan's revolutionary new way of working, believing that high quality came at a high cost. The idea that the rigorous application of a small number of fundamental, continuous-improvement concepts would dramatically improve both quality and cost seemed unreasonable. As a result, over the ensuing years many American companies and hundreds of thousands of jobs disappeared. Many books and articles were written during this period about the end of the "American era" (Dowd, 2007; Vogel, 1979). After suffering substantial commercial and social pain, America eventually adopted these profoundly more productive ways of working, as did the rest of the world. Now, every significant manufacturing company uses some version of TQM continuous-improvement principles.

This approach has been so effective that today, low cost and high quality are the entrance requirements for most new products. Now companies must increasingly move to a broader definition of customer value. The innovation economy demands high quality and low cost, but it also requires that we deliver new products and services with more convenience, features, personalization, design, and user control, among many other ways to create additional customer value. It also demands that we take the same approach to the other aspects of the enterprise: manufacturing, distribution, marketing, human resources, financial systems, legal services, and information technology.

IV. THE WAY WE WORK IS THE MOST IMPORTANT INNOVATION[11]

Can we, like Deming and Ohno, achieve dramatically better results by developing and using more productive ways of working? At SRI we strongly believe this is possible through the comprehensive application of the fundamentals of innovation, which are not widely known or applied today. Although interest in the topic of innovation is great, the field of innovation concepts and best practices is still in its infancy. It is like the

[10] "Toyota Motor Corporation," *New York Times*, July 15, 2010 http://topics.nytimes.com/top/news/business/companies/toyota_motor_corporation/index.html

[11] From C. R. Carlson, who says about SRI's innovation practices, "The way we work is our most important innovation."

discipline of TQM before Deming and Ohno codified and popularized the core ideas (Shewhart, 1931).

To test the maturity of innovation understanding, ask seasoned executives for the definition of innovation. You will typically be told that it is about creativity, teamwork, intellectual property, novel ideas, or entrepreneurship. These definitions are incomplete and lead to confusion and inefficiency. Every enterprise requires a comprehensive "innovation playbook," and few have one today.

A complete definition for innovation is: "The creation and delivery of new customer value in the marketplace. Innovations are sustainable only if they provide sufficient enterprise value to allow for their continued production."[12] A product or service may be clever or creative, but unless customers in the marketplace use it, it is not an innovation. As a dramatic example, consider that the US Patent Office has so far issued more than 4,000 patents for mousetraps (Hope, 1996). Yet only about 20 of those thousands of patents have ever made money.[13] The others may represent clever, creative ideas but they are not innovations. Unless an enterprise obtains sufficient value for producing the product or service, it rapidly disappears and ceases to be an innovation.[14]

Innovations can be small and transitory, like Motorola's flat RAZR phone, or large and long-lasting, like Thomas Edison's light bulb, or the computer mouse with interactive computing developed by Douglas Engelbart (Nielson, 2006)[15] or the Internet. Whatever the size of an innovation, individually or cumulatively, it is possible that over time the accumulation of innovations can create enormous new customer value.

Consider Ford's Model-T compared to today's automobiles. Both are still means of transportation, but today's automobiles include a tremendous number of both small and large innovations. It took many tens of thousands of small innovations to achieve the remarkable quality, durability, and reliability of today's automobiles. In addition, today's automobiles can include many major innovations, such as air conditioning, AM-FM-satellite radio, airbags, seatbelts, GPS-guided navigation systems, communication systems,[16] and pollution controls. And, unlike the Model-T, which came only in black, the choices now include a rainbow of colors.

Outputs, Not Just Inputs: It is important to focus efforts on outputs—innovations—and not confuse them with inputs. Concepts like entrepreneurship, creativity, collaboration, intellectual property, and business skills are all inputs that can lead to new innovations. The goal is not entrepreneurship per se (the set of skills, attitudes, and behaviors that can help a person be more successful at innovation), it is innovation itself.

Using the wrong words to describe innovation can cause confusion, limit success, and discourage people from participating fully. For example, after I gave a talk on innovation to a large group of academics, a department head of mechanical engineering said to me, "That talk changed my life " (Carlson, 2008). When asked why, he said, "Because I have been asked to teach entrepreneurship, and I don't feel like an entrepreneur—that is not who I am; it is not my identity. Teaching entrepreneurship has always made me feel uncomfortable. But I am passionate about innovation. That is why I obtained my Ph.D., became a professor, and agreed to be a department head. It is also why I love teaching students, so that they can become innovators and make positive contributions too. Now I realize that I can teach these

12 This definition is slightly different from the one given in Carlson and Wilmot, 2006: 6, but the meaning is essentially the same.

13 See http://uh.edu/engines/epi1163.htm

14 "Sufficient value" means that the producers can either recoup their ongoing investments or they can find a way to have the endeavor subsidized. The airline business is an industry that, cumulatively, has generated negative financial returns over its history. It survives only because of government subsidies and because individuals continue to invest in it. Wikipedia is another interesting case. Here, the subsidy comes from people's time, which they provide to make a subject they are interested in available to the world. Open-source software is still another. There are many ways an innovation can be sustainable other than through financial profit for a company. Obviously, most innovations are transitory but some, like the wheel, can last a very long time. An innovation's significance is clearly a function of its longevity, the number of people for whom it delivers value, and the total financial value it creates. That is why the wheel is often thought of as one of the world's greatest innovations, along with language and cooking. In modern times, many believe that the Internet is the most important innovation. See http://en.wikipedia.org/wiki/Timeline_of_historic_inventions for an interesting list of the world's greatest innovations.

15 See also http://www.sri.com/about/history/nielson_book.html

16 Like *OnStar* by General Motors

"A complete definition for innovation is: The creation and delivery of new customer value in the marketplace. Innovations are sustainable only if they provide sufficient enterprise value to allow for their continued production"

courses with enthusiasm using the new understanding you gave us today." This attitude is very common among technical professionals, whether in a university, a company, or government.

Innovative Understanding: Many thousands of executives, technical managers, academics, and government officials from around the world have come to SRI International's headquarters in Menlo Park, California, to participate in a program called the SRI *Five Disciplines of Innovation™*.[17] The program begins by asking participants to write answers to a series of questions, including "What are the definitions of innovation, customer value, and a value proposition?" These are among *the* most basic concepts in any business. Remarkably, only about 20% of the participants can reasonably answer these questions when the program begins. By not having a common, accurate language for the most basic concepts of innovation, their strategic decisions and day-to-day interactions are often confused and inefficient. Clearly, these basic ideas are not widely taught or understood.[18]

Fundamentals of Innovation: Many authors have contributed excellent ideas about how to think about and improve innovative success (Drucker, 1993; Christiansen, 1997; Moore, 2002, and Porter, 1998). Important concepts include "crossing the chasm," "open innovation", "industrial clusters", and many more. These concepts, however, are best applied after the fundamentals of innovation are in place. In the book, *Innovation: The Five Disciplines for Creating What Customers Want*, a family of fundamental "disciplines" are described, which are used by SRI and many of its partners (Carlson and Wilmot, 2006: 20). SRI's five disciplines are:

1. Important customer and market needs
2. Value creation
3. Innovation champions
4. Innovation teams
5. Organizational alignment

Each of these disciplines describes a set of concepts and best practices that increase the probability of innovative success. These disciplines have proven to work through extensive application and experimentation over many decades.[19] They provide a focus on customers' needs, both internal and external, and they offer a common language, concepts, tools, and processes for rapidly amplifying the process of value creation. SRI believes that these five disciplines are effectively multiplicative. If an enterprise rates a "zero" in any one, the probability of success is also effectively zero. If several are implemented poorly, then the enterprise's innovative potential is significantly reduced.

Value Creation: It is not possible to describe all five disciplines here. Rather, this section describes elements of "value creation" to illustrate several basic

[17] SRI International http://www.sri.com

[18] The innovation economy requires changes in the educational curriculum too, such as a more comprehensive understanding of innovation. This includes fundamental business concepts and a global perspective. Today's graduates must be able to write clearly and give compelling presentations, which have become even more important. Finally, they must have the human skills and values needed for productive, multidisciplinary collaboration.

[19] One of the most thoughtful contributors to our understanding of the process of knowledge creation is Douglas Engelbart, the inventor of the computer mouse and the foundations of personal computing at SRI in 1967. (Carlson and Wilmot, 2006: 169, and http://dougengelbart.org/)

principles. The section entitled "Case Study—SRI's Journey" will briefly describe the other four disciplines and their application.

Developing a new innovation is not an event; it is a *process* that requires the creation of new knowledge—*value creation*. It is a process, as illustrated in Figure 1, where new knowledge at A is applied to address a customer need at B to create a new product or service. From B to C the enterprise generates profit, but eventually the product or service becomes obsolete and the value creation process must be repeated.

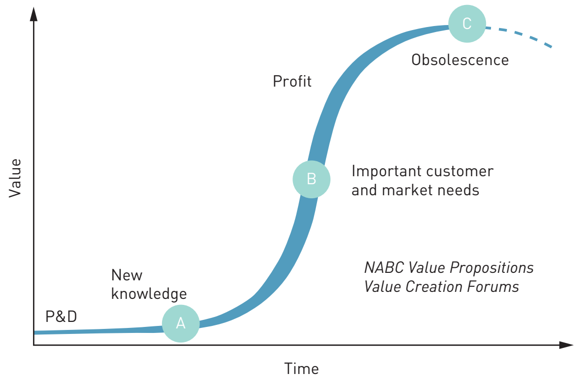
All innovations require connecting A to B. This process is very hard and it takes great skill, effort, and considerable time to develop a compelling, high-value solution. Often this process is called the "Valley of Death" because it is so difficult to understand and navigate (Taylor et al., 2008). At every step, the most efficient and effective practices should be used.

Because connecting A to B is common to *all* innovations, any advance that makes the process faster and more successful is itself a major innovation—a meta-innovation. It is for this reason that we say, *"The way we work is the most important innovation."* Below are several examples of concepts, tools, and processes that greatly increase innovative efficiency and the likelihood of success.

Value Propositions: Developing a new innovation starts by answering four fundamental questions, which define the proposed innovation's *value proposition*:

1. What is the important customer and market *Need,* not one that is just interesting to you?
2. What is the unique, compelling new *Approach* to address this need?
3. What are the specific, quantitative *Benefits per cost* (i.e., customer value[20]) of that approach?

Figure 1. Value creation is a process where new knowledge at **A** and an important customer and market need at **B** converge to create an innovation, which generates enterprise profit from **B** to **C**. At some point the product lifecycle is complete, the product is obsolete, and it is necessary to create a new, higher-value product or service. The role of R&D is to provide new knowledge to address important customer and market needs. Innovation tools and processes help facilitate value creation, such as NABC Value Propositions and Value Creation Forums.



4. Why are those benefits per cost superior to the *Competition* and alternatives?

These four questions define what SRI calls an "NABC Value Proposition" (i.e., Need, Approach, Benefits per costs, and Competition, Carlson and Wilmot, 2006: 85). *Every* new innovation must answer at least these four questions: they are the absolute minimum for any proposed new innovation. Focusing on these four questions, rather than starting by trying to write a 300-page report, saves enormous amounts of time because in the beginning little is known about the customer or the market; seldom have the best ideas and partners for the approach been identified; and typically little is known about the competition and alternatives to the new idea. Thus, there can be little to no understanding of the possible benefits per costs.

Hypothesis-Driven Innovation: A value proposition is begun with an initial *hypothesis*.

[20] Customer value is defined two ways: Financial Value = Benefits – Costs and Perceptual Value = Benefits/Costs. See ibid, Carlson and Wilmot p. 79

This can be an observation about a market trend, or a paradigm shift in technology, or any number of other insights. This is the proverbial "light bulb" switching on. But no matter how clever a flash of insight, it must be expected that this first hypothesis will be wrong. Indeed, if it is a significant new innovation, the final product or service will be very different from what was imagined at the beginning.

SRI has found that *none* of its major innovations ended up where they started.[21] If it is a major new innovation, the reason for this, as has just been stated, is that so little is known at the start. Rather, a hypothesis is made, data is gathered and synthesized, a new hypothesis is developed, and then more data is gathered and synthesized to create still another hypothesis. This iterative process continues until there are solid answers for all elements of the value proposition. It takes unrelenting iteration to get to a reasonably good, quantitative value proposition. In essence, would-be innovators should, "fail fast and fail often to succeed early."[22] New iterations should be daily or weekly at the start.

The NABC method focuses innovators on the most fundamental questions first, which are *very* hard to answer. It saves enormous time and effort that, unfortunately, is often spent by untrained, would-be innovators on useless activities. Once the NABC Value Proposition is developed, one can move forward, and efficiently create a more detailed innovation plan. The NABC approach applies to all functions in an enterprise, whether in R&D, finance, HR, branding, or new product development. That is because if you have customers, whether outside or inside the enterprise, you can always create more value for them. Even for the most basic research, one should be able to answer these four basic questions.[23] Only after these four questions are answered can a more complete innovation plan be efficiently developed.

"Big-A" presentations: If you hear many presentations, you are probably frustrated by how difficult it is to understand whether they are describing anything of importance. Mostly they are focused on their "approach" with little useful information about the market, customers, and competition. They proclaim that the market is huge, people will love the product, and that there are no competitors or alternatives—ever. But there is *always* competition. We call these "Big-A" presentations, as in nAbc. They are all about the approach—i.e., the person's great new idea. To a potential funder or partner, Big-A presentations have essentially no value. All four questions—NABC—must be compellingly, quantitatively answered to have a meaningful conversation about the potential value of a new idea. Big-A presentations create enormous confusion and inefficiency—waste.

Value Creation Forums: An important process for speeding up value creation and avoiding Big-A presentations is to tap into the "genius of the team". At SRI, these meetings are called Value Creation Forums. [24] The objective is to rapidly improve innovative ideas and to create compelling Value Propositions. Two guiding principles make the meetings most productive. First, everyone stands up and presents: no bench-sitters allowed. Each person gives an NABC Value Proposition about their important project.[25] They present for five to ten minutes and, when time is up, they must stop. The presentations are short so that the presenters focus on the fundamentals, which are very hard to answer. Second, the presenter's teammates then critique the presentation to reinforce what worked and suggest how it can be

[21] Conversation with Norman Winarsky, vice president of ventures and licensing for SRI International and his colleague Vince Endres at the Sarnoff Corporation (a wholly owned subsidiary of SRI), 2010

[22] There are many versions of this idea (Kelley, Littman and Peters, 2001).

[23] These four questions are almost identical to those asked by the United States investment agency DARPA (Defense Advanced Research Projects Agency) in their requests for proposals.

[24] Carlson and Wilmot, 2006: 101, where Value Creation Forums are called "Watering Holes," a colorful term that does not fully translate into some languages. Value Creation Forums are used across SRI International to develop new innovations ranging from new cancer drugs to new Web-based software companies.

[25] A better format is an "Elevator Pitch," which starts with a "Hook" to gain interest, the NABC Value Proposition, and ends with a "Close" to end the presentation and ask for a specific action, such as a date for a full meeting (Carlson and Wilmot, 2006: 128).

improved.[26] The presenter listens carefully without responding to the input: corrections can be made later to save the group's time.[27] This approach has proven effective in corporate, academic, and governmental settings because they all require that the fundamentals of an NABC Value Proposition be addressed for every new initiative.

Experience shows that after three or four Value Creation Forums, with a partner helping in between Value Creation Forums, the improvements made are impressive. Note, however, that if the innovation is significant, many dozens of meetings are required before the answers needed are obtained. Value Creation Forums allow for the rapid sharing of ideas while allowing each participant to be a role model for their teammates. In addition, these meetings tap into participant's natural competitiveness, which incentivizes them to rapidly improve each presentation.

Why a Playbook? Innovation concepts and best practices constitute a "playbook" for employees.[28] Without a playbook it is almost impossible to systematically succeed. Consider, as an analogy, football or soccer players and their playbooks. No professional team can win without them. They describe a set of specific plays, what each football player will do, and how each player will coordinate their efforts with their teammates as the play unfolds. These plays are practiced over and over until everyone fully understands them and they can be precisely executed. Professional coaches help the players understand the plays and apply "best practices" to speed up learning. Of course the playbook will change depending on the players available, the competition, and the environmental conditions. Once the game starts, the players must adapt and modify their plays in response to what the competition does. In addition, there are broken plays and it is often necessary to improvise. But because they have practiced diligently over many years, players have a portfolio of possible "improvisations" that are understood by their teammates and that have a reasonable chance for success in different situations.

Most people do not think about innovation this way, but having a playbook focuses everyone's efforts, keeps the team moving in the right direction, and coordinates the team's efforts. Innovation is very much a "contact" team sport, where players must execute their roles professionally and efficiently. And, yes, every new innovation's competition is on the move too, and an innovation team must continuously adapt and improvise. But if the team is prepared and open to adaptation, it is much more likely that the changes needed to succeed will occur. Very few organizations use a comprehensive playbook of innovation concepts and best practices, but the ones that do are often impressive. [29] These practices represent a major source of competitive advantage for such companies.

An Innovation Laboratory: With its industry, academic, and government partners, SRI has been responsible for numerous world-changing innovations, which have created many tens of billions of dollars of new economic value.[30] SRI has been studying the best innovators around the world, teaming with them on projects, and inventing new innovation concepts and best practices.

SRI is unique in that it is both a major innovation practitioner and an "innovation best practices laboratory", where the concepts listed above have been developed and tested with thousands of colleagues, both inside and outside of SRI. SRI has discarded practices that are ineffective and kept those

[26] An additional option is to have each presenter share one new "innovation best practice" of value to the team, so that the team can learn additional concepts about innovation.

[27] It is best to have someone taking notes for the presenter as the feedback is given.

[28] This concept also comes from personal discussions with Pallab Chatterjee and the author.

[29] For example, examples include SRI International, Medtronics, IDEO, Toyota, and P&G.

[30] SRI's innovations with its partners include the computer mouse and modern human-computer interface; electronic banking; the United States high-definition TV standard; treatments for cancer and infectious disease; minimally invasive robotic-assisted surgery; computerized speech recognition; the world's first virtual personal assistant (e.g., *Siri Inc.*); and much more.

> "Very few organizations use a comprehensive playbook of innovation concepts and best practices, but the ones that do are often impressive"

that work. Most of the ideas tried were not effective because they were either too complex or not valuable to staff. They may have sounded good in an academic setting, but when applied by professionals working to solve real-world problems, they were not. Over and over, SRI has learned that it is the core, fundamental concepts that make the biggest difference in terms of sustained innovative success. SRI has also discovered, by working with dozens of leading companies around the world, that few enterprises even try to seriously apply them. The concepts seem easy to understand, but that does not mean they are; they are not. They can *only* be understood through thoughtful, vigorous, and steadfast application.

Changing Role of Management: The innovation economy requires that management redefine elements of their jobs. Consider first, as an extreme case, Henry Ford. His initial management approach was severely top-down. He wanted to make essentially all significant decisions about his company. He even had detectives monitoring his managers; and if any of them deviated from his orders, they were fired.[31] Consider as another extreme example, academic management, which is in many ways still all bottom-up, controlled by tenured faculty (Garrett and Davies, 2010). Academic management is rightly acknowledged to be an extremely difficult, often frustrating task.[32]

The advantage of top-down management is that decisions can be made quickly. The advantage of bottom-up management is that it allows for a multiplicity of new ideas. But neither is ideal. In the innovation economy all top-down is increasingly uninformed and all bottom-up is increasingly irrelevant. Finding the "sweet spot"—the right balance between top- down and bottom-up—has always been a difficult task (Brafman and Beckstrom, 2006). But, as a rule, the sweet spot for management influence has been moving down in the organization because of the rapidly changing dynamics of the innovation economy.[33] Only front-line employees are in daily contact with customers, markets, and technologies and are able to rapidly make accurate decisions. By contrast, senior managers who have worked their way up the corporate ladder are mostly familiar with a previous time's customer needs, market dynamics, competitors, and technologies. Consider that only 20 years ago, the World Wide Web was just beginning as was 2G mobile communications. Over this 20-year period, computing power has improved by roughly ten thousand times at the same price. Applications like Google, Facebook, and Craigslist were almost unimaginable just a few short decades ago.

In the innovation economy Henry Ford's style of management is increasingly archaic because one person cannot possibly learn enough, fast enough about customers, markets, competition, and technology. It is not smart enough.

Just as top-down alone is increasingly out of date, so is all bottom-up. There are exceptions, but many of the most important

[31] See http://www.whatsbestnext.com/2010/02/an-example-of-bad-management/

[32] Garrett and Davies, 2010: 70, "The management of creative professionals starts and ends with encouraging, supporting, and incentivizing achievement". This quote was given to Garrett and Davies by C. R Carlson in 2010.

[33] This also makes it harder for some more traditional managers, who want control.

opportunities today require multidisciplinary teams to create meaningful solutions. The apparent paradox for many managers is how to create an enterprise where there is sufficient freedom for invention, yet enough structure to capture the ideas generated and turn them into valuable innovations. Letting staff go off in a hundred different directions does not produce value; it produces organizational chaos. Programs that emphasize "inspiration rooms"[34] or "innovation centers," or the trappings of creativity, such as pool tables, funny hats, play dough, and LEGO blocks, are often, by themselves, misguided.

At the other extreme, in the innovation economy, academia's style of management is also increasingly archaic because it does not support collaboration within a disciplined innovation structure. It is not smart enough either.

What is required is an organizational architecture, like the one described below, for the disciplined incubation of new high-value innovations. It requires new organizational structures that better exploit the best features of top-down and bottom-up. These new innovation structures complement the more traditional structures, such as TQM and stage-gate management systems, which remain effective for incremental innovations.[35] But TQM or stage-gate structures alone are inadequate.

Benefits for Employees: Innovation skills are important to a company's staff. People with the ability to innovate are among the rarest people in the world: they are always in demand. Experience shows that when professionals gain these innovative skills, they become more successful while helping their enterprises to be more successful. The quality of their R&D and innovation initiatives improves; their ability for productive collaboration with colleagues and partners increases; and a conceptual framework is created for more rapid learning and continuous improvement. Having these skills allows for greater career achievement and professional growth, which means that enterprises supporting this kind of environment are preferred by the best employees.

V. CASE STUDY—THE SRI JOURNEY

The ideas described above have had a transformational effect on SRI, which has had a storied history in Silicon Valley. Stanford University's creation of SRI 65 years ago was one of the seminal events in the early formation of Silicon Valley, along with Hewlett-Packard. Most professionals probably use several SRI innovations every day, whether it is the computer mouse, multiple computer windows, high-definition television, electronic banking, computerized speech recognition (through Nuance Communications), automated mail sorting, and minimally invasive surgery (through Intuitive Surgical).

SRI is an innovation enterprise: that's all it does. SRI has worked in nearly half the countries in the world and in all major technological areas. SRI has pioneered management concepts now invoked widely, such as "SWOT" analysis and "open innovation". Since its founding, all of SRI's major initiatives have been based on open innovation, because they were all completed with great partners. In spite of these enormous achievements, by 2000 SRI had stopped growing. The innovation concepts and best practices SRI had pioneered up to that point were no longer enough. With the emergence of the innovation economy around that same year, a more comprehensive innovative approach

[34] See, for example, http://www.theinspirationroom.com.au/who-is

[35] Note, other innovation management approaches are often called "Stage-Gate" and "Funnels." See Wikipedia, *Stage-Gate*, http://en.wikipedia.org/wiki/Stage-Gate_model. From SRI's experience with many international companies, these approaches often have limited to no success in creating major new innovations.

was required. That was the year SRI began to rigorously apply the *Five Disciplines of Innovation*.[36]

Since 2000, SRI has had a dramatic turnaround with double-digit growth, a cadre of staff doing R&D to solve more important problems, and a much more valuable venture and licensing pipeline. In 2010 alone, SRI had a major cancer drug approved by the US Food and Drug Administration for T-cell lymphoma, a terrible cancer for which there was previously no good therapeutic. In addition, one of SRI's spin-off companies, Siri, was bought at a premium by Apple Computer,[37] even though it was only 18 months old at the time. Siri is the world's first practical computer assistant, a major advance in personal computing. In the future it may be seen as rivaling in significance the development of the computer mouse more than 40 years ago.

SRI's Innovation Architecture: SRI applies the *Five Disciplines of Innovation* to all aspects of its business: R&D, new product development, venture formation, and all corporate functions. Use of a common language based on customer value has elevated cross-divisional communication to a new level. It allows incremental innovations to be developed more efficiently, and puts a more productive focus on larger, multidisciplinary initiatives, which are required to solve important problems.

SRI uses a family of Value Creation Forums to create new innovations. Across SRI are market-focused forums, which match SRI's strategic focus areas, such as cyber security, infectious disease, intelligent computer systems, education technology, and clean energy. These Value Creation Forums are focused on generating compelling Value Propositions; they do not fund new R&D. Until a good Value Proposition is developed, it is a waste of resources to spend money on technology.

SRI's market-focused forums are each organized and facilitated by an expert in the specific market area. The forums are given limited resources each year [38] to be spent on consultants, market studies and reports, customer and partner visits, and product designs and simulations. These forums come and go depending on market conditions and SRI's ability to contribute. SRI has two other ongoing Value Creation Forums, one for R&D investments and one for commercialization activities. These forums have much larger investment resources and there are senior managers who run the meetings, act as mentors to new potential innovators, and negotiate major transactions.

SRI is a transparent enterprise where anyone can talk to anyone else without permission. For example, Value Creation Forums are posted on SRI's internal website. All staff can come to a forum, but it is understood that these are value creation meetings where everyone is expected to contribute. Just sitting and watching is not enough. With freedom comes responsibility.

Comprehensive Application: SRI uses its common innovation language wherever possible. For example, there is an "SRI Card," a wallet-sized plastic up card that describes the company's mission, vision, values, and many of its innovation practices. SRI aspires to be the "premier independent source of high-value innovations". The *Five Disciplines of Innovation* are described during the hiring process and they are on SRI's website. New employees are more formally introduced to the disciplines of innovation at orientation. SRI's professional development focus is on how staff can use the *Five Disciplines of Innovation* effectively, along with how they support SRI's vision and business objectives.

[36] C. R. Carlson became CEO of SRI in 1999. Previously he was a vice president of business development and ventures at SRI's wholly owned subsidiary, the Sarnoff Corporation.

[37] Interestingly, at the start of Apple Computer, Steve Jobs licensed the computer mouse from SRI.

[38] Value Creation Forums are given tens of thousands of dollars annually.

The CEO personally holds an innovation workshop for all new employees to indicate the importance of these practices. He often has lunch with members of the staff and asks them about issues needing improvement, their work, and their Value Propositions.

SRI promotes an "abundance" mentality, not one based on scarcity. But it makes it clear that it is only a world of abundance if staff have the required innovative skills and are able to apply them effectively. Appropriate incentives are in place to focus outwardly on customer and market needs—that is, on value created.

All business presentations at SRI use the NABC format. They drive investments, speed up iteration, and minimize the need to compare "apples with oranges." SRI works hard to keep presentations short: one-page proposals, 15-slide presentations, etc. SRI is focused on outcomes—real value to its customers. Because of its common language, concepts, and tools, staff members understand each other more quickly, input is more consistent, and new insights can be incorporated more easily. There is much less confusion about what staff and management are agreeing to do and why.

Important Market and Customer Needs: In the innovation economy, we must aspire to work on important customer and market needs, not just those that are interesting to us. Interesting problems are quickly overrun by others in the innovation economy. Important customer and market needs allow for the creation of significant customer value and they also motivate and attract the best staff.

As described, Value Creation Forums at SRI are all run using the same basic language, concepts and tools. Beyond the basics, the expectations and presentations required for specific tasks are quite different. For example, the metrics for success from a new venture are dramatically different from those expected from R&D. If SRI is going to start a new venture, it must be worth at least several hundred million dollars to be of interest. This is not an arbitrary objective. Among other reasons, this threshold is required in Silicon Valley because, if it is not met, it is extremely hard to acquire the best management team and venture partners. Other activities have metrics that are appropriate to the task, such as those for new R&D centers. These metrics allow staff to decide more easily whether an initiative will have value for customers and SRI. It is surprisingly rare for management to have to say no. Rather, a team proposing a new idea soon realizes whether the threshold goal can be met and, if it cannot, the idea often goes away.

Innovation Champions: Without a person who is passionately committed to making a new innovation happen, it will fail. The first question SRI asks about any investment, proposal, or project is, "Does someone really want to do this?" Will someone commit to success, no excuses, and agree to follow the *Five Disciplines of Innovation*? SRI has a saying, "No champion; no project; no exception." If the idea is a good one and SRI has no champion, they do not start serious work until one is found.

This is SRI's approach throughout the organization, top to bottom. Champions are born with many of the traits needed for success, but they must also be nurtured and cultivated. Training in innovation begins in the technical divisions and progresses to corporate venues. High-value innovation is about achievement. That is what motivates people and gets them to work day and night. You can never force people to work this hard unless they are passionate about their work.

Every major innovative initiative must be built around that fundamental human need, which champions possess.

SRI focuses on its innovation playbook to help staff achieve their goals. SRI is in the highly competitive Silicon Valley—if the playbook does not work, the staff will not use it. Even so, it takes a great deal of management effort and time before new staff fully understand what SRI aspires to achieve with its innovation concepts and best practices, how to apply them, and why they will be valuable to their careers.

Innovation Teams: In the innovation economy, an enterprise must team with the best to maximize its chances for success. Even large companies rarely have all the best resources. Although almost every company will claim that it abstains from the "not invented here" (NIH) syndrome, the truth is that almost all suffer badly from the disease. Because they do not normally assemble the best teams, they are effectively hoping that their competition fails to do so as well. Obviously, if their competition does assemble a crackerjack team, they may be defeated in the marketplace.

Forming teams is hard. It is a project that must be actively managed. It takes training, support, encouragement, and appropriate rewards to have staff create powerful, productive teams. To overcome the frictional costs of team formation, major goals are required—e.g., important customer and market needs. The cost of putting together a high-powered team is otherwise not justified. An advantage of working on important customer and market needs is that there is an abundance of psychic rewards that can be distributed throughout the team.

Organizational Alignment: Organizational alignment starts with senior management making a commitment to ensure that the enterprise will be a market leader and that they will achieve this by delivering the highest customer value in the minimum time and at the minimum cost. It means putting in place the structures, metrics, rewards, staff, and support to satisfy the *Five Disciplines of Innovation.* It means removing obstacles to innovation. A common example is barriers to staff. When they need to ask a vice president in a different division a question, they are required to get permission from several levels of management. In addition to slowing down the process of value creation, that sends exactly the wrong message to staff about the enterprise's commitment to the rapid creation of high-value innovations.

Achieving the goal of becoming an innovation enterprise must be at least a five-year initiative. Progress is relatively slow at first but then momentum builds: you will not go back. Build forward motion through early adapters; focus on achievement and impact; demonstrate value; and create internal ambassadors. As the saying goes, "Lead with the best to push the rest."[39] Involve everyone at the strategic level; but deeply involving everyone is not possible. Make receipt of funding contingent on using the *Five Disciplines of Innovation* to the extent possible—this shows seriousness. The innovation agenda will not be taken seriously if it is too marginal. Make Innovation concepts and best practices a core business process in as many venues as possible. Focus on the fundamentals: the greater the market and customer attention and connection, the better the results.

No organization can ever achieve perfection, but every organization can strive to get better through a serious commitment to continuous improvement. SRI strongly believes in asking every enterprise activity to improve some aspect of their function each

[39] This is a common saying of Dennis Beatrice, who is vice president of the Policy Division at SRI. He also contributed many ideas to this section.

year. SRI is not close to where it wants to be, but each year it gets better. Success takes substantial time, but even modest progress creates significant returns.

VI. CONCLUSIONS

We are in the innovation economy. There has never been a better time for creating major new innovations: it is potentially a time of abundance and unprecedented prosperity. But it is also the most challenging time in the history of innovation, with technological improvements in most fields occurring at rapid, exponential rates and with global competition increasing equally dramatically. This dynamism will not stop. These driving forces will accelerate as billions of people in the developing world move from poverty and low-cost manufacturing to prosperity and the creation of new, high-value innovations.

Our innovative performance today is, overall, poor. Few companies have comprehensive innovation playbooks for staff with an organizational architecture that drives innovative success. Both are essential for survival today. Creating an innovative enterprise starts with commitment by senior management and then by putting the fundamentals of innovation in place. Once these fundamentals are established, it is possible to add other innovation concepts to further develop the enterprise's innovative sophistication. The fundamentals are not hard to understand, but they are extremely hard to practice. The only way to really learn them is through repeated application. Few make the effort but those that do often excel.

Experience shows that large improvements in innovative performance are possible. Even a ten percent improvement would make a significant contribution to the profitability of most enterprises. In many cases, improvements have gone well beyond that. Having a deep understanding of innovation is beneficial to staff too. Professionals today need new skills based on a comprehensive understanding of the innovative processes that lead to success. Those who have these skills can prosper: those without them will increasingly fail. Enterprises that help their employees obtain these skills have an advantage in attracting and keeping the best talent.

The innovation economy gives us the opportunity to create abundance through the application of a discipline of innovation. To thrive we must use innovation concepts and best practices throughout our enterprises and more generally throughout industry, academia, and government. The way we work is *the* most important innovation. Even small improvements in our collective ability to innovate would, over time, have a huge positive effect on the world's prosperity, environmental sustainability, and security.

REFERENCES

http://en.wikipedia.org/wiki/PayPal
http://en.wikipedia.org/wiki/Population_decline
http://en.wikipedia.org/wiki/Stage-Gate_model
http://en.wikipedia.org/wiki/Timeline_of_historic_inventionshttp://www.sri.com/about/history/nielson_book.html
http://mt.kargo.com/v/News/HowChinaWillChange/?KSID=3189d3546687c862a6eebeb2eaf0ef7bhttps://www.cia.gov/library/publications/the-world-factbook/
http://uh.edu/engines/epi1163.htm
http://www.commerce.gov/news/press-releases/2010/07/13/locke-announces-national-advisory-council-innovation-and entrepreneurhttp://topics.nytimes.com/top/news/business/companies/toyota_motor_corporation/index.html
http://www.sri.comhttp://dougengelbart.org/).
http://www.theinspirationroom.com.au/who-is
http://www.whatsbestnext.com/2010/02/an-example-of-bad-management/
BRAFMAN, O. and R. BECKSTROM (2006), *The Starfish and the Spider*, Portfolio Hardcover.
CARLSON, C. R. (2008), "The Imperative for Including Innovation in a Technical Education", Keynote at the Kern Family Foundation Annual Meeting at Thunderbird University in Arizona, January 7.
CARLSON, C. R., and J. SCHAUFELD (2011), "Technical Education in the Innovation Economy", to be published in *Shaping Our World,* John Wiley and Sons.
CARLSON, C. R., and W. W. WILMOT (2006), *Innovation: The Five Disciplines for Creating What Customers Want*, New York: Crown-Random House.
CHRISTENSEN, C**.** M. (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Boston: Harvard Business School Press.
DEMING, W. E. (1986), *Out of the Crisis*, The MIT Press.
DOWD, A. (2007), "Three Centuries of American Declinism", *Real Clear Politics*, August 27.
DRUCKER, P. (1993), *The Practice of Management,* New York: Harper & Collins.
FOSTER, R., and S. KAPLAN (2001), *Creative Destruction: Why Companies that Are Built to Last Underperform the Market—How to Successfully Transform Them*, New York: Doubleday-Currency.
FRIEDMAN, G. (2009),*The Next 100 Years: A Forecast for the 21st Century*, Doubleday.
FRIEDMAN, T. (2010), *The World is Flat: A Brief History of the Twenty-First Century*, Farrar, Straus and Giroux.
GARRETT, G., and G. DAVIES (2010), *Herding Cats: Being Advice to Aspiring Academic and Research Leaders*, Devon: Triarchy Press.
HOPE, J. (1996), "A Better Mousetrap", *American Heritage Magazine*, October, pp. 90-97 and http://www.americanheritage.com/articles/magazine/ah/1996/6/1996_6_90.shtml
HOSAKA, T. (2010), "China overtakes Japan in 2Q as No. 2 Economy", *MyWay*, August 16, http://apnews.myway.com/article/20100816/D9HKI1E00.html
KAPOR, A. (2010), "Urban Greatness Awaits Good Governance", *New York Times*, May 20.
KELLEY, T., J. LITTMAN and T. PETERS (2001), *The Art of Innovation,* New York: Crown Business.
KURZWEIL, R. (2005), *The Singularity is Near: When Humans Transcend Biology*, Viking Press, September 22, 2005.
MANDEL, Michael (2009), "America's Innovation Shortfall", *Bloomberg BusinessWeek*, June 3.
MARKOFF, J. (2005), "It's Moore's Law but Another Had the Idea First", *New York Times,* April 18.
MITCHELL, L., "Moving Innovations to Markets", *Kauffman Foundation*, http://www.kauffman.org/advancing-innovation/moving-innovations-to-market.aspx
MOORE, G. A. (2002), *Crossing the Chasm*, Harper Paperbacks, August 20.
NIELSON, D. (2006), *A Heritage of Innovation: SRI's First Half Century*.
OHNO, T. (1988), *Toyota Production System: Beyond Large-Scale Production*, Productivity Press.
PORTER, M. E. (1998), *Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press.
RIDLEY, M. (2010), *The Rational Optimist*, Harper Collins.
SHEWHART, W. A. (1931), *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand Company.
STONE, Douglas (2008), "Winning the New Product Innovation Game", *Corp!*, December 4.
TALEB, N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House.
TAYLOR, E., *et al.* (2008), "Encouraging Industry-University Partnerships", *Kaufman Foundation*, http://www.kauffman.org/uploadedFiles/EAC_UIP_report_v4.pdf
VOGEL, E. F. (1979), *Japan as Number One: Lessons for America*, Replica Books; republished in 2001.
WADHWA, V. (2005), "Ah out that Engineering Gap", *Business Week*, December 13.

BBVA

Designing Radical Innovation

Harry West
Continuum

THE INNOVATION IMPERATIVE

The world is changing: customers, consumers, constituents—the users of our products and services—are now in control. Users are more discerning, they have more choices and are more comfortable exercising these choices. Users are demanding that products and services work for them in the fullest sense: that their total experience is as good as it can be. Organizations, companies and even governments are finding that they have to address the demands of their users, and do so with an unaccustomed urgency. If they do not, their users will exercise their choices and change brand, change behavior, or change their government. This is the innovation imperative; we cannot rely on momentum in the market to protect us. Customers, consumers and constituents have learned how easy it is to change, and so we have to make sure that we are always providing the very best experience to stay ahead of their expectations.

What are the forces driving this change? The prime movers are communication technology and the rapid growth in new markets. Communication has fueled the growth in democracy from the fall of the Berlin wall, to the ascendancy of democracy in Latin America, to customer reviews on Amazon.com. When people can see alternatives they make choices. And following the frontline of democratic change has been the baggage train of growing consumer markets as citizens became empowered and resources were redistributed more equitably and more productively. At the trivial end of the democracy scale, communication has made it possible for citizens to choose not only their government, but also the winners on American Idol, Britain's Got Talent, or Tú sí que vales. No longer are our stars chosen by record industry executives or an elite panel of judges, or in a long, painful process of working their way up. Today we get to choose them real time, live. We are becoming used to the idea of exercising individual choice, expressing dissatisfaction with the status quo, and expecting an immediate response.

Communication also fosters technical parity; it is easier to learn, to track what others are doing, and to copy. MIT has made available the best technical lectures in the world to everyone with a broadband connection, free. In China we are seeing knock-offs preceding the original brands to the market. Today, there is little difference in the chips in our computers, in the engines in our cars, or in the ingredients of many of the foods we eat. Since the ingredients are the same, it is what we do with them that differentiates the experience of the product or service. We have to broaden what we deliver to be a more complete solution, or risk becoming just a supplier of commodity components in the global business ecosystem.

And, of course, our increasing ability to communicate has had a direct impact on how people spend their money. As a species we like to communicate. We get pleasure from talking with others and sharing the stories of our lives, and over the millennia we have learned of new things through word of mouth. Social networking tools have expanded this process over the last couple of decades. These tools greatly increase the network of advisers who tell us what we could be experiencing, how to get it, and the best offers out there. An increasing part of our monthly budget is spent on cell phone service, broadband, smart phones and other communication technologies. More of our products and services incorporate communication as an integral part of their experience; for example automobiles are now communication centers with satellite radio, cell phones, navigation systems, etc. And when we buy music or books, seeing how others have reviewed them is an integral and valuable part of the buying experience.

In parallel with increasing communication has been the growth of two extraordinary new markets, Gen Y and Emerging Markets. Gen Y—broadly speaking people born between 1980 and 2000, have grown up in a time of mostly peace and prosperity, of pervasive communication and of low-cost computing technology—for them this is the norm. They are a large population, 80 million in the US and over 100 million in Europe. The old models of music, automobiles and banking do not work for them; they have different expectations. Not all young people in the US aspire to owning a car the way their parents did, and when they do buy one they care as much about communication and entertainment as horsepower. In addition to leading the adoption of many new specific digital and social products, they are also driving the integration of digital tools and social networking into everything else—the integration of the physical and the virtual. To hark back to an earlier time, it was not the plastic product that was revolutionary; it was the plastic in everything. Each generation establishes its own norms of material expectations and lifestyle, and Gen Y is now defining the new middle class.

Globally the *growth* of the middle class is primarily an emerging market phenomenon. Each year about 70 million people enter the middle class in emerging markets—about the population of France. As these populations become consumer societies they are adopting many of the same expectations as North America and Western Europe, albeit with local variations. And even amongst people who are not yet middle class we can see the effect of new communication technology changing behavior. In Peru, one of the poorest countries in Latin America, frequent use of the internet extends down to level C. And ask teenagers there what would they would do if they were given a Sol and they will probably tell you they would head to a nearby internet kiosk to chat with their friends or check out what is happening on hi5.com. The Gen Y population in Latin America is approximately 200 million, and in India and China even larger with approximately

400 million each. These populations dwarf those in North America and Western Europe. They are so large that the preferences of Gen Y and emerging markets are no longer the tail of the consumer dog, they are the dog and North America and Western Europe are the tail. When the dog wags its tail established markets are shaken. We have seen this already in consumer electronics; companies based in what just a few decades ago were emerging markets have taken the lead, Sony (Japan) eclipsed Zenith (US), Samsung (Korea) is in the process of eclipsing Sony, and LG (Korea) has set its sights on overtaking Samsung. Meanwhile, fast followers in China are pulling out into the passing lane.

DESIGNING RADICAL INNOVATION

What do we mean by designing radical innovation? To innovate is to change something established by introducing new products or services, new processes, or new ideas. By radical innovation we mean that the change is fundamental; it is not the same thing only shinier, it is a complete reset of the customers' experience or a completely new business model. By design we do not mean a drawing to show how something looks; we mean the purpose and the plan for the innovation. Design is the deliberate process for purposefully creating the new.

Our motivation for designing radical innovation is not that we want to be radical for its own sake, but that the situation calls for it—the innovation imperative is radical. There are now so many points of connection between smart people, new technologies, and ambitious businesses that it is inevitable that innovation will happen somewhere in the business ecosystem. Increasingly, consumers, customers and constituents are taking control and innovating for themselves[1].

Our challenge is to help our own companies to benefit from change rather than become a victim of it. How does a company deliberately see what's next, recognize how they can profit from it, and then move their organization to lead the change?

This is difficult to do. For many businesses, innovation has not been a central part of their strategy. Successful businesses can be based on the continued exploitation of some earlier innovation, and focus on narrower goals of cutting costs, improving quality, building distribution, and promoting sales; generally tightening control over operations. It is only when there is a disruption, such as a new or reinvigorated competitor, that innovation becomes important again. The trouble is that by then the organization may have lost the ability to innovate. The trick is to learn to innovate before you are cornered.

Innovation requires loosening control. The essence of innovation is that you do not know in advance what is going to emerge, so the organization cannot define and control the innovation process in the same way it does everything else. The only rules are rules of thumb, and the organization has to put its trust in an unknown outcome. To make matters even more challenging for the business leader, it takes time to find the business value in innovation, and then time to communicate the innovation plan to the larger organization, and then more time to execute it. Radical innovation has a longer-term cadence that does not play well with quarterly reporting cycles

Successful innovation requires the careful evolution of new ideas. If we try to innovate too quickly we end up with ideas that merely reflect the current needs of consumers. The ideas will be liked today, but if they do not stake out new territory

[1] Eric Von Hippel, "Democratizing Innovation," The MIT Press, 2006.

in the minds of consumers they will not be distinct enough to be owned in the future. For many organizations it is difficult to get the pace right. They have learned to make decisions at the end of every meeting. This is great for efficiency and showing clear leadership, but not so good for nurturing new ideas. On the other hand, going too slowly is just as bad. If the project is not on a 'critical path' it becomes a second priority, part time project that risks suffocating any last traces of creative potential. The trick lies in pacing innovation so that it fits into that magic creative window of opportunity with enough time to allow ideas to percolate and enough urgency to keep the team motivated.

Innovation is counter to many of the business processes instituted in the drive towards six sigma quality, and the people who are most capable of catalyzing innovation in a company may be outside its traditional structure. To learn how to innovate, many companies are working with outside groups experienced at thinking differently—a stimulus to jolt the system into a new state. Often, these outside groups include designers, whose processes and perspectives offer distinct advantages.

Why is the design process good for managing innovation? Designers are innately motivated by the "new"; that is why they are designers. They have a complementary mindset to the "control" prevalent in most large companies. More than any other discipline, designers are naturally comfortable with abductive[2] processes in which creation precedes analysis—the knowing is in the leaping. Their "new" muscle is well developed; throughout their training they have been challenged to create ideas that are new and to communicate them to others so that they can then be critiqued and evaluated. Over the last fifteen years designers have also learned to develop stronger customer empathy through ethnography and other techniques for immersion in the lives of the people they are designing for. As our economy migrates from being about just products to one that trades in broader experiences, designers are able to use their customer empathy to deliberately craft the emotional journey that is an essential part of experience innovation. And in interdisciplinary design, firms and department designers have learned to collaborate with like-minded engineers and business people and develop strong critical thinking skills. In parallel with the development of the innovation capabilities of individual designers, we have also developed a well honed "design process" to provide a structure for the leap into the unknown. This process is not difficult to learn, but the social and personal dynamic that supports creativity, customer empathy and collaboration is difficult to enculturate.

INNOVATION DESIGN PROCESS

There are four essential elements in the process of how radical innovation is successfully designed and deployed. We call these elements of the process rather than stages because they occur simultaneously, not necessarily sequentially—innovation is an on-going integrated process. You can think of these elements as four different spaces in the organization.

1. Executive Space

The first element is at the executive level of the organization. It is how you acknowledge the need for change, accept that you do not know yet what that change will be, and prepare the organization to embark on a process of learning what to do and then executing in accordance with what it learns.

[2] Roger Martin, "The Design of Business: Why Design Thinking is the Next Competitive Advantage", Harvard Business School Press, 2009

Challenges at this level are that there needs to be a continued focus on executing the current strategy at the same time as the organization is learning what to do next, and often these will be in conflict. As the great American novelist F. Scott Fitzgerald wrote, "The true test of a first-rate mind is the ability to hold two contradictory ideas at the same time." In navigating change you need to manage both what is right for today and what will be right for tomorrow, and continue to remain open to completely new ideas that are not even in the framework of your current thinking.

You are asking your organization to commit to embarking on a journey without knowing the final destination.

2. Consumers, Customers, Constituents Space

The second element is how you immerse your organization in the lives of its consumers, customers and constituents. The organization is to search for new solutions, not inwardly as experts, but through the lens of its users. Your team will be conducting research as if they were anthropologists visiting an unknown tribe. The purpose of the research is for the team to be able to empathize with users and to help the team to create from the users' point of view (see sidebar on customer research).

Challenges at this level are that users probably do not know what they will want in the future so you cannot abrogate your creative responsibility to them; your job is to create for them. Ideas that come out of this research are provisional and their validity is not knowable until they have been envisioned and modeled in a prototype form that can be tested.

You are asking people with the least formal power in the organization, not executives or experts, but users, to direct its destination.

3. Project Space

The third element is building a team to lead the process, evaluating it to make sure that the innovation is right, and deploying it so that it can be executed by the larger organization. A radical innovation team will comprise a range of people with complementary skills, disciplines, and mindsets. It should include people who understand well the constraints that the business will have to work within, and should also include people who do not see any constraints at all. The team needs a strong leader who has the confidence to listen as well as to make decisions; the leader should listen carefully to what the team is feeling even when they are obviously wrong. High-performing teams take advantage of communication technology to work remotely, but are collocated for much of the time in a dedicated project room.

A challenge at this level is that you are asking a team of people to completely focus on one thing—the future—for a long time. In that time they will become distanced from the main part of the organization that they will need eventually in order to execute the innovation.

You are asking a scouting team to make a long commitment to find the way for the main organization.

4. The Prototype Space

The fourth element is the prototype space where you model and evaluate a range of ideas to learn, iterate and refine the innovation. Great ideas with small flaws fail; details matter. The prototype space is how you get the idea right. It is also how you help the rest of the organization to understand what it is that you are going to do, and enlist their support.

Learning from customers

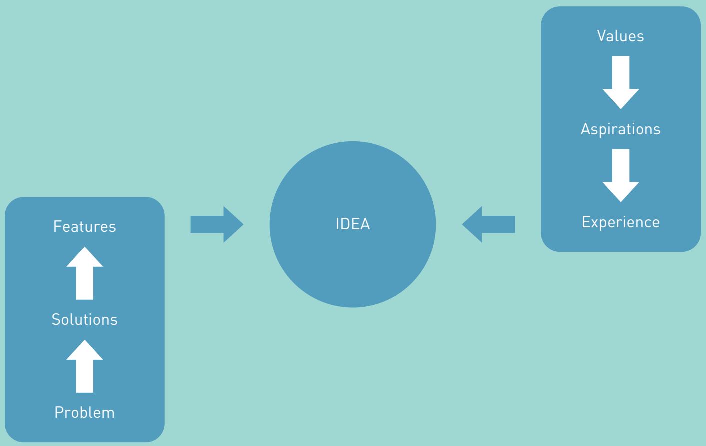
The inspiration for innovation can come directly from watching or listening to people, but often we have to reflect deeply to find the meaning in the research that has been hidden from us in its simplicity or familiarity.[3]

We immerse ourselves in the lives of people—consumers, customers, constituents—in an open-ended way, acknowledging that not only do we not know the answer; we do not even know the question. We must watch and listen to people to learn from them how our organization can help. We need to unlearn our vocabulary, and learn again the words that people use. We want to uncover the issues that concern people, unaided and unbiased by our preconceptions.

To do this we typically spend one to four hours with about seven people per segment. We watch them in their context; their home, work, hospital, on the street, in a store, or whatever is the relevant context. We engage them in a conversation about their lives.

Of course, such a small sample cannot be used to infer proportions in the population, but we are looking for issues that are important to most people, and you are unlikely not to uncover these issues if you have a sufficiently in-depth conversation. Remember Newton: the story goes that an apple fell on his head as he was sitting under a tree, and from that stimulus he came up with the laws of gravity and a theory to explain the motion of the planets. He did not sit under a thousand apple trees; he used that time to think instead.

So how do we think about what we learn from customers? We look for two types of information. The problems they have; can we think of a way to solve their problems? And their values that make them want to participate in a category in the first place: can we create an experience that speaks to those values? A great idea is one that combines features that solve problems in an experience that speaks to people's values.



[3] Apologies to Wittgenstein

Swiffer

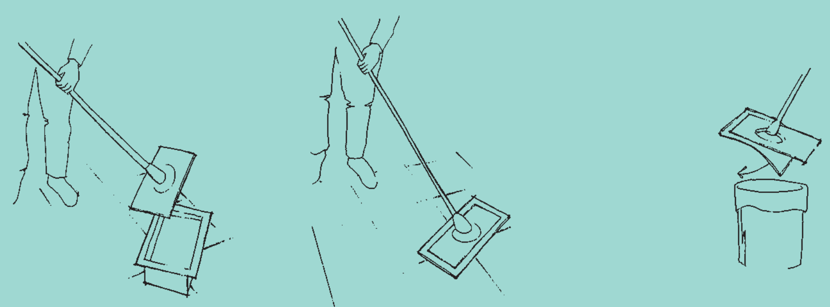
Rethinking the everyday

In 1994 Procter & Gamble was looking to grow through innovation. Craig Wynett, Director of Corporate New Ventures asked, "Is there a better way to clean a floor?" The answer to that question not only reinvented how we clean floors, but also changed how we think about the role of design in driving business innovation.

A joint team from Continuum, Northlich Stolley, and P&G set out as design anthropologists to watch how people cleaned their kitchen floors. We noticed that most people swept their floors before they mopped and they had to assemble a system of products to get the job done. We noticed that mops worked mostly by the adhesion of dirt to the mop and that people seemed to be spending almost as much time rinsing their mop as they did cleaning the floor. We noticed that people wore old clothes when they were cleaning because it was a dirty job, and we could not help but notice that sometimes people had cleaned their floors before we got there, because they did not want us to see them cleaning a dirty floor. People were concerned about the cleanliness of their homes.

Out of watching those 18 women in Boston and Cincinnati (yes, they were all women, tsk, tsk) emerged an idea that we affectionately called "a diaper wipe on a stick" and that in consumer testing we called "Fast Clean". When we first showed consumers the concept for "Fast Clean" they did not like it because they did not believe that it would clean well and they thought it would be expensive. We made a working prototype, an experiential model, to communicate the idea. Once people had experienced using "Fast Clean" they were sold.

It took several years to go from the first sketches and prototypes to launching what we know as Swiffer, but it is now one of the most successful new products that P&G has ever launched with sales of over $500 million per year. Swiffer succeeded because it is not just a new mop or a new broom, but a completely new and seamless home cleaning experience. It solves the technical problems people encountered cleaning their floor, by making the process quicker and cleaner for them, and it spoke to people's value for the cleanliness of their home.



Insulet

Launching a new business

Insulin pumps are the most effective tool for managing diabetes because they closely mimic the function of a pancreas—delivering a constant stream of insulin to the body. The Insulet Omnipod is the first discreet, disposable, human insulin pump and has launched a new category.

Prior to Insulet the stakeholders in diabetes care—the patient, doctor, and insurance company—often advocated multiple daily injections instead of a pump. Patients struggled with technical issues: pumps were complicated, confusing, and uncomfortable at the injection site. Insurance companies were wary of the up-front costs associated with buying pumps, which were often abandoned. Doctors found pumps too complicated for patients to use effectively.

But the biggest barrier was an emotional one: unlike injections, a pump and its tubes broadcast to the world that the user had diabetes, and technical constraints prevented the user from showering, swimming or sleeping while wearing it.

Insulet, a small VC funded start-up, partnered with Continuum to systematically solve these problems. Together, we developed a pump system that was not just a medical solution for patients, but a human solution for people.

The OmniPod is the first wearable insulin pump that attaches directly to your skin. It is small, low cost, and disposable; after three days you simply replace it with a new one. It reduces pain by automating the insertion of the cannula into your skin. It is easy to use: instead of buttons on the pump, it is controlled wirelessly using a handheld device that looks like a cell phone. The wearable pump provides complete freedom—you can shower, swim, and sleep wearing it. And it is discreet—there are no tubes; managing your blood sugar looks like you are texting a friend. It is a completely new healthcare experience.

Each one of these breakthroughs is a revolution in diabetes care. Bringing them together was a business breakthrough. Doctors prescribe Omni-pod because they have confidence in its simplicity and efficacy. Insurance companies support it because its costs are distributed over a 2-3 year period. Users want it because it gives them back their freedom and privacy.

Working as Insulet's user research, product design and engineering development team, Continuum partnered with Insulet for 3 years through initial 510k clearances. Insulet launched the OmniPod and completed an initial public offering of $116 million.

Spence Diamonds

Helping people to make a lifelong decision

Choosing an engagement ring is a big decision. Spence Diamonds is the largest diamond retailer in Canada and their business depends on helping people through this process.

Talking with customers in-store and in their homes we heard their anxiety about cost, style, and about making such an important decision. After all, it is not just a diamond; it is a commitment.

We choreographed a new sales approach that is more self-service so that customers can manage the pace themselves, reducing anxiety and freeing up staff to engage later in the process when their expertise is most needed. We designed breaks into even this last part of the customer journey for our happy but anxious couples to talk in private; sometimes good service means stepping back. There are also moments when the brand should step forward to share its expertise and to take some responsibility in the decision. Based on what we heard from our customers we instituted a new, simpler language supported by professional tools to give customers the serious information they need to make the right "investment" decision. This new sales process was executed through the design of the retail environment, crafting of collateral and fixtures, and training for employees.

The net result is a better way for Spence Diamonds to help their customers that goes beyond traditional one-on-one customer service, enabling Spence Diamonds to continue to expand their business and to support their customers in the right way.

A challenge at this level is allocating the resources to build experiential models that will enable the learning and refinement needed before committing to launch. It is sometimes easier for an organization to commit resources to a specific idea even if it is not right, than to fund the open-ended exploration of the right idea.

You are asking your scouting team to show the rest of the organization where they will be going by building a facsimile of it in detail.

These four essential elements in managing radical innovation are linked by the project team, but cannot be the responsibility of the team alone. Without executive support they will become disconnected from the mission of the company.

> "*Radical innovation design* asks first, how can we make the experience of our customers' lives better? And then only second, how can we use this to drive business innovation and support the needs of our company?"

DESIGNING EXPERIENCES

In the process and examples just discussed we have turned convention on its head. *Radical innovation design* asks first, how can we make the experience of our customers' lives better? And then only second, how can we use this to drive business innovation and support the needs of our company? This is counter to a conventional approach that starts with the opportunity for the firm and then looks for ways to make it wanted by customers. Even among sophisticated companies that are espousing customer centricity, it is the day-to-day language spoken by executives and written in presentations that demonstrate whether or not they are truly thinking about the future from their customers' point of view, or just trying to be more sophisticated in how they exploit them. To be clear, we understand the needs of business and our motivation is not altruistic or philanthropic, but we have seen how companies that benefit consumers the most succeed the most. The way to business success is through a relentless focus on the good of the consumer.

The change we are going through is a self-reinforcing cycle. As consumers, customers and constituents experience a better product or service their expectations ratchet up and they continue to look for what is even better. Furthermore, people have learned to transfer their expectations from one category to another: *If it is so easy for me to buy music online, then why can't I buy shoes the same way? Don't tell me reasons why it is difficult; I don't care that music is digital and shoes are physical; the cost of shipping is your problem; I have no respect for the former hierarchy in the shoe industry and will go with whoever solves this problem for me*. Today, Zappos.com is a major US shoe retailer with sales of over $1billion a year. A big part of their success goes beyond just the shoes themselves: Zappos offers free shipping, free returns, and is constantly looking for ways to interact more with their customers through their website, by phone, or through social media.

As a result they have become widely known for their great customer service and have a very loyal customer base. Zappos has created what for many people is a better shoe buying experience.

Putting customers, consumers and constituents at the center means we consider the total experience of the user. Consumers improvise their experiences out of the products and services available to them; they make choices from available alternatives and substitute one for another. By designing better and more complete experiences we give the consumer bigger building blocks to construct their lives, which makes it simpler to construct richer experiences. We reduce the need for the consumers to improvise, and make it less necessary for them to consider different options. Taking responsibility for more of the total experience of the user is in the best interest of the firm because in a world of technical parity products are easily copied, but a seamless customer experience is more difficult to imitate.

Experiences are more personal and emotional than products. Designing experiences means we need to be attuned to local nuance. As corporations developed global businesses, they moved from locally designed products to globally distributed ones. We are now in the process of transforming again to "globally local experiences" that are customized to the specific needs of each group of local consumers.

We are nothing but the experiences of our lives. We want these experiences to be as good as they can be, and we want them to be validated by others. As users—customers, consumers and constituents—we are becoming accustomed to being in control and are demanding the right experiences; the simplest experience, the most reassuring experience, the richest experience. Some experiences are seemingly trivial, like cleaning the floor; some are life changing, like managing diabetes; and some are emotionally charged, like buying an engagement ring; but together these experiences add up to the whole of our material lives. Each part of the human experience deserves our respect and should be designed with care. As a result of advances in communication technology and increasing globalization it is now possible to think more carefully than ever before about each element of our experience and to amortize the design cost over ever larger numbers of users. As designers we are privileged to be able to use our talents to help people in this way.

> "By designing better and more complete experiences we give the consumer bigger building blocks to construct their lives, which makes it simpler to construct richer experiences"

Innovation: Behind the Buzzword

Pascal Sobol
IDEO

What is the difference between two world-class cities like Sydney and New Orleans? The former has an impressive concert hall that hosts memorable performances, while the latter is virtually synonymous with great music. In other words, Sydney *hosts* wonderful "cultural experiences," whereas New Orleans *cultivates* them. Because New Orleans maintains an artistic mind-set—and so many of its residents and visitors identify with its creative purpose—musicians from far-flung places flock there, eager to play, experiment, and challenge the status quo.

When it comes to innovation, similar parallels can be drawn in the business world. Some companies routinely hire qualified R&D staff who produce worthwhile products or services—and help the organization stay competitive in doing so. But other firms let creativity permeate their corporate culture, encouraging all employees, across silos and hierarchies to take creative approaches to planning and problem-solving. In other words, some companies rely on the 'creative' experts hired into their innovation departments to drive their innovation attempts, while others have made innovation a corporate priority across all departments.

So what? Isn't being "innovative" just one of many goals a company pursues? And doesn't it go without saying that some are better at it than others? And yet, while many business leaders claim that they want to be more innovative in order to stay competitive—and increasingly believe that innovation is essential to an organization's survival—they rarely pause to consider to which level this actually needs to shape their organization as a whole.

In order to truly make innovation a competitive advantage that a business can bank on, innovation needs to evolve beyond being an activity a few are engaged in and become a state of mind across the entire organization. Let us take this opportunity to examine why an organizational culture of innovation is so critical today, and how you as a business leader can foster one.

WHY BE INNOVATIVE?

Every other year, IBM asks several thousand CEOs worldwide to comment on the state of the economy and to identify their greatest business challenges going forward. Organizational "complexity" currently tops the list. The latest study (May 2010) also shows a remarkable shift in how executives think they should address this complexity. For the first time ever, CEOs said that the most important leadership quality to possess today is "creativity."

Reading the details of the study, one realizes that these CEOs do not just mean creativity in the sense of good ideas for new products and offerings. It goes far deeper. These executives refer to a clear demand to creatively approach every aspect of their organizations so they can adapt to rapid and sustained market change. They want to experiment constantly with new business models and technologies, as well as to shape and continuously reshape their relationships with customers. These CEOs need system-level creative thinking to come up with innovative means to address system-level complexity.

Based on its interviews, the study's authors singled out three ways to achieve this: embody creative leadership, reinvent customer relationships, and build operating dexterity.

In other words, even business leaders who previously relied on the "Sydney" approach to innovation now see the value of bringing a "New Orleans" mind-set to their organizations.

A decade ago, it was sufficient to rethink a company's offerings at regular intervals. Today it is critical to evaluate the whole organization—including its products and services, its structure, and its relationships with customers and employees—on a continuous basis.

SHIFT HAPPENS

Many cutting-edge businesses are already putting these three ideas into action. Creative leadership at Zappos, for example, involves offering a considerable sum of money to any new employee who decides to actually *leave* the company, to ensure that staffers who stay *really* want to be there. Meanwhile, Nike is proving that reshaping customer relationships can be a serious business opportunity—its "Nike ID" co-creation platform, which allows users to design their own sneakers within predefined limits broke the $100 million turnover barrier earlier this year. And Apple has been teaching the world about operating dexterity for years: it repeatedly demonstrates an uncanny ability to branch out into new markets by rethinking them on a systemic level and developing highly desirable offers that disrupt incumbent offerings.

Embody creative leadership. "Creative leaders invite disruptive innovation, encourage others to drop outdated approaches and take balanced risks", the authors noted. "They are open-minded and inventive in expanding their management and communication styles, particularly to engage with a new generation of employees, partners and customers."

Reinvent customer relationships. "In a massively interconnected world, CEOs prioritize customer intimacy as never before. Globalization, combined with dramatic increases in the availability of information, has exponentially expanded customers' options", the authors wrote. "CEOs said that ongoing engagement and co-creation with customers produce differentiation. They consider the information explosion to be their greatest opportunity in developing deep customer insights."

Build operating dexterity. "CEOs are revamping their operations to stay ready to act when opportunities or challenges arise", the researchers observed. "They simplify and sometimes mask complexity that is within their control and help customers do the same. Flexible cost structures and partnering capabilities allow them to rapidly scale up or down."

(IBM CEO study, p.10, May 2010)

The motivations for business leaders to creatively innovate is not only to have competitive offerings out in the market, but to make sure the organization as a whole remains able to quickly respond to a business environment, which is changing ever more rapidly on every level.

HOW TO MAKE AN ORGANIZATION MORE INNOVATIVE?

It might not be obvious how business leaders can go about bringing out system-level creativity within their organizations in order to address the system-level complexity they have been observing. There is however, a breed of companies who have experienced the marriage of creative approach and strategic consultancy for years, our company being one of them. So some of the answers are already out there. Let's have a closer look at a few of them.

What the CEOs in the IBM study are essentially describing is the need for a more design-driven approach to organizing and running their companies, one that is insight-led and richer in experimentation.

Designers are adept at reinventing the way they work along with whatever they are creating. Strategic innovation firms like IDEO are taking on a broader array of challenges for clients in increasingly diverse industries, from developing a customer experience at government service centers to designing an early seed investment firm from scratch. Our approach, now known as "design thinking", evolved over the past decade by applying a human-centered design process to myriad projects. Once valued exclusively within the "creative profession", design thinking is now taught by major business schools to answer complex strategic challenges in nearly every industry.

CHANGE BY DESIGN

Design thinking, inspired by real-world observations, is rooted in human behavior and needs. As such, applying design thinking allows business leaders to base their decisions on reality rather than on theory or assumptions. It also enables CEOs to make strategic output tangible much earlier in the process than traditional management approaches. As design is at the core of this approach, strategies tend to be expressed as brand prototypes, hypothetical advertising and use scenarios instead of pie charts and bar graphs. This makes the outcome of a strategy project much more communicable and easier to grasp within the organization.

With the IBM study rating creativity as the top leadership quality, it is clearly the right time to look at how design thinking can be applied to make whole organizations more innovative. So, taking our cue from the world's CEOs let's look a little deeper into each of the high level opportunity areas that they have outlined. Based on the study's structure, we can highlight a number of recurring themes that we at IDEO have observed in our daily conversations about making organizations more innovative.

1. EMBODY CREATIVE LEADERSHIP: *Purpose, Not Vision*

In order to harness innovative forces within the company, it is easy to assume that senior management's primary responsibility is to project a strong vision of where the company is going. After all, the classical business paradigm positions the CEO as a "visionary". Unfortunately, the idea that the boss always knows best doesn't ring so true anymore. In times of extreme complexity, it is unrealistic to expect C-level management to have command of all of the answers, despite how thoroughly they may be briefed by team members.

Yet in most organizations today ideas are still more or less expected to come from the top. The CEOs in the IBM study, however, seem to throw their hands up and say, "Listen guys, that's just not how it works anymore. With everything else we are responsible for, we can't be the ones monitoring every technology trend, every social phenomenon and every business play that is relevant to our company."

This admission that leaders do not have all the answers is a sign of strength rather than weakness: it invites all employees into the dialogue about where the company is and should be going.

The most progressive companies are more sharply defined by a unifying purpose than a leader's vision. A "vision" implies top-down culture and suggests that management sets the course while the rest of the organization merely executes that vision. However, innovative organizations are increasingly aware that there is no one "right" way to do things and that no single person knows enough to have a perfect master plan.

Leaders as Curators

Instead, innovative leaders see themselves much more in the role of someone curating an exhibition. They allow for—and demand—experimentation from their employees within the boundaries of an overall framework that they as curators provide**.** In turn, they make sure that everyone internally and externally is aware of the organization's purpose—i.e., what it stands for over all, in order to make sure that experimentation is compatible with company values.

IKEA, for instance, has been carefully built around the purpose that its founder Kamprad set out thirty years ago: "We have decided, once and for all, to side with the many. What is good for our customers is also, in the long run, good for us", he said. He wanted to "create a better everyday life for the majority of people". Note that this is not a vision about the future of flat-pack furniture or where the brand can go. This is a mission statement of almost political dimensions. Every business decision can be measured by it, yet it leaves plenty of space for the organization to develop into areas that Kamprad was unable to foresee at the time.

Top-Down and Bottom-Up

In order for innovation to enter everyone's consciousness throughout an organization, innovative thinking needs to be encouraged and rewarded from the top down. At the same time, leadership needs to provide the space for good ideas to grow and blossom from the bottom up. This may sound simple, but it is contrary to how most organizations operate today. The predominant corporate mind-set is one of hierarchy and of avoiding risks *at all costs*.

Employees of most organizations tend to believe that if something is "not their job", they don't need to take any responsibility for it. Yet innovation is too important to be left solely to designated Innovation Units. Although specialized personnel can prove useful, no one group will ever be able to live up to the creative potential of everyone's brains combined. The specialist's job is not necessarily to have the answer, but rather to locate and assemble the people who know most about the task at hand and to give them permission and the necessary space to be creative in a focused way.

Innovation is Everyone's Job

When a chain of private hospitals in the US asked IDEO to help it improve its patient

Finding real-world inspiration by going through a patient experience

experience, our team stepped in to manage a co-creation process. All the ideas that we generated and prototyped came from doctors, nurses, patients, and families. When we invited and expected everyone to contribute their best thinking to the project and gave them credit for its outcome, the ideas came pouring out of them, which meant we could concentrate on "curating" the process, or setting up some general guidelines and making sure that everyone was aligned on the purpose of the project. In doing so, our team managed to make innovation a mind-set among all hospital staff.

IDEO isn't alone in its design-thinking approach. Procter and Gamble's former CEO A.G. Lafley recognized the need for spreading innovative thinking throughout the firm much earlier than many of his competitors. In his book *The Game Changer*, he concluded, "Innovation is everyone's job." When Lafley revamped P&G's innovation infrastructure in the early 2000s, he underscored his expectation that everyone on staff was an innovator by creating the P&G Gym. "The Gym" is a laboratory environment designed to rapidly develop ideas and initiatives so that they either excel or fail quickly. Any internal team can check in to the Gym for innovation guidance and support to take projects to the next level. This enables P&G to support the best ideas, no matter where they originate inside the company.

Look Beyond What is Measurable

Most organizations try to eliminate ambiguity anywhere in the development process. In order to avoid ambiguity, they designate quantitative data as the main information currency. This is a good thing when pursuing goals like quality control or regulation compliance. The problem is that to develop truly groundbreaking ideas, innovators have to look beyond what is measurable. By the time a need can be quantified in surveys, it has passed from the latent to the explicit stage and chances are that it has been addressed by some competitor. In other words, once people can clearly articulate what it is that they need, the solution is just around the corner. Insights that are critical in creating breakthrough innovation cannot be expressed primarily in numbers. As innovators, we need to trade in a different currency and leadership needs to be able to accept that currency.

In order to enable innovation teams to arrive at game-changing solutions, leaders must alleviate employees' fears that they will be shot down in boardrooms if they lack sufficient quantitative data to make their cases. Instead, leaders should encourage teams to work more like detectives than statisticians and proceed without the perceived safety nets of pie charts and percentages. Sometimes all it takes is one seemingly small insight to lead to a breakthrough idea.

Just like detectives discovering clues they can't necessarily explain right away, innovators find insights which are hard to evaluate at first. But processed correctly, the value will be relatively quick to determine in what is ideally an ongoing dialogue with potential customers. The quality of these insights serves as the currency in highly innovative organizations.

Embrace Ambiguity

A few years ago, on a project for Bank of America, an IDEO team was developing innovative offers for mothers. The market had been flat, and customer turnover significant (the base was expensive to maintain). When the team met with a few dozen moms, two insights emerged: First, the act of saving was emotionally pleasing, even if the saved amount was very small. Second, some women in the target group had the curious habit of rounding up their utility bills paid by check. Nobody knew what these observations would lead to, and we initially presented them as just that: observations. Over the course of the project, however, it turned out that these insights established fertile ground for cultivating ideas.

One idea was to combine a checking and savings account in order to make the act of saving unconscious and game-like: Every transaction on the checking account gets rounded up, and the difference is deposited into the savings account. This evolved into Bank of America's "Keep the Change" program, which has turned out to be an attractive and easily understandable proposition for many customers. The program allowed the bank to divert part of the marketing dollars spent on regaining customers to a matching scheme aimed at retaining them. Since "Keep the Change" was introduced, more than 12 million customers have signed up and collectively saved more than $3 billion.

The "Keep the Change" program was based entirely on the careful observation of a few individuals. If the IDEO team had worked under the pressure of having to quantify the value of the insights at every turn, the idea would never have left the boardroom.

Analyzing and interpreting insights to make them strategically valuable

2. REINVENT CUSTOMER RELATIONSHIPS

Procter & Gamble's Lafley not only fostered a culture of innovation within his company, but also placed considerable emphasis on customer relationships. He demanded that everyone in his organization get in touch with the end consumer. Lafley was known to show up for home visits in order to set an example and stay informed. He realized that, in order to make good decisions, his teams needed to have their intuition grounded in the real world. He understood the importance of customer relationships not only for marketing purposes, but also for developing relevant offers.

A decade later, many business leaders recognize that the consumer needs to play a central role in the innovation process. However, opinions seem to increasingly differ about how to best achieve this user involvement, especially when it comes to input early in the process. Some companies still rely on surveys and quantitative studies. Others have branched out and started collaborating with consumers via online tools, which now provide a plethora of communication channels that companies can use to get the consumer's point of view represented early on in the process.

From "What can we do?" to "What should we do?"

It seems all too obvious to state that innovation should be driven through the benefits delivered (to consumers?), not through the tools that allow us to deliver them. Yet in many of today's organizations, innovation is still anchored solely within R&D departments. Traditionally, most research and development teams have a technology-centric worldview. While looking to technology for innovation is a valid way to generate ideas, it is an inherently limited one. When holding a particular piece of new technology in hand and looking for suitable applications, it is hard to avoid the syndrome of "if I am a hammer, the world looks full of nails".

The most radical examples of innovation may feature new technology, but they were designed from deep consumer insights. Take the poster child of innovation, the Apple iPod. The iPod did not pioneer any particular technology; it reinvented an experience.

Generating a breadth of ideas quickly through a series of brainstorm sessions

The device was based on the insight that consumers craved a sleek system that let them easily discover, purchase, store, and enjoy music on the go. Various technologies were then combined to create the system. Yet it is unlikely that Apple would have arrived at the same solution by looking at a cutting-edge hard drive at the time and wondering "So, what could we do with this?"

Today, engineers, coders and scientists are capable of creating just about anything. The question has shifted from the R&D-oriented "What *can* we do?" to the customer-focused "What *should* we do?" The latter is, of course, a question that organizations must answer as a whole.

Companies should be wary of the so-called "technology push trap." It is always seductive to use new technologies as a starting place when looking for innovative offers. But the danger lies in losing sight of human need by narrowly focusing on technology. We have found that successful innovation most often occurs when engineers, user specialists, and business experts jointly shape an initiative from the start. As an interdisciplinary team, together they can ensure that any solution proposed will be desirable, technically feasible, and financially viable.

Find Real-World Inspiration

There is a common dilemma for companies who are leading in the innovation game: For them, it is not as obvious where to look for new ideas as it is for followers. The competition cannot provide much guidance, as that would essentially mean looking backward. Many innovation pioneers in diverse fields have figured out that the key is to be inspired by real-world behaviors and attitudes. This allows them to find and act on opportunities for a step change in their markets before anybody else can disrupt them. How they detect these opportunities is a crucial part of their continuing success.

Every large organization has a market research department and is capable of conducting surveys and evaluating concepts with end consumers. The reality we see, however, is that while many organizations consider themselves knowledgeable about their consumers, they are all too often drowning in a sea of data, without gaining any insight.

Planning real-world inspiration gathering as a series of open-ended, inspiring exercises with tools like ethnography and analogous experience audits is usually the best way to approach an innovation challenge.

Inspire first, quantify later—much later

In our experience research tools are often used in a misguided fashion. It has become customary to apply quantitative methods early on in the process, when what is needed is in fact not numbers but inspiration. The sense of security that quantitative methods provide is clear: in order to convince anyone internally, employees feel they need to deliver statistically relevant numbers backed up by representative samples of their target audience. The need to verify through numbers is understandable, but it stifles inspiration and should be suppressed until much later in the process. Quantitative methods are great for evaluating sufficiently developed concepts, but they don't inspire new thinking early on. On the contrary, applied at the beginning of the process they tend to make the search for innovation slow, expensive and ineffective. Quantitative studies are not designed to allow for unexpected results, because the realm from which interviewees choose their answers has been pre-determined when designing the study.

Individuals Inspire Big Changes

Inspiration can usually be gained much more efficiently through an ethnographic approach, in which very few individuals are observed, but are interviewed in great depth. This approach is much less expensive and time-consuming and much more likely to point innovators in interesting directions.

One particular IDEO project provides an illustrative example. While working for a credit-card provider, our design team interviewed a woman who considered herself to be an impulsive shopper and had racked up considerable credit-card debt. When our research team asked her whether she was happy with the credit card services her bank offered her, she replied, "Sure". A quantitative survey would have stopped there and noted that the user was satisfied with the offer. But when we asked her to take us through her online purchasing routine, we learned that she kept her credit card

Interviewing ATM users about their financial perceptions and observing their behaviors

BBVA ATM – humanizing self-service banking

in a block of ice in her freezer, instead of her wallet, to help her resist using it. When quizzed about her behavior, she proceeded to tell our team at length how she did not trust herself with making quick purchasing decisions, so she forced herself to wait until the block of ice had thawed to free the card until making the purchase. This gave her added time to consider and reconsider her decision.

This example illustrates how a design-thinking approach can unearth a deeper truth than traditional methods: The "quantifiable" survey would have showed a customer who was satisfied with her credit card. The human-centered observations delved more deeply into how the customer actually used the card. The fact that she was trying to modify her self-perceived negative behavior patterns provided design insight and inspiration. She is unlikely to be unique in this need, which begs the question as to what kind of products banks may want to offer that help credit-card users overcome their own shopaholic behavior. Whether the need actually does exist widely can be tested within weeks of generating new ideas and shaping them into concepts.

Target Beyond the Target Group

Our designers start every project by spending considerable time with a few end customers to understand not only what they say and do, but also what they think and feel. When we set out to "unlock the commercial possibilities of the self-service channel" for BBVA, for example, we observed and interviewed people who use automated teller machines on a regular basis. We did the same research with people who did not use ATMs. Uncovering the limitations of a current offering requires understanding not just the target group, but also the "extreme" users, or people whose behaviors are *in some relevant way extreme*.

For instance when searching for inspiration while engaged in a project with the flip-flop brand Havaianas, our teams spoke not only to Brazilian beach goers and urban European fashionistas. In addition, they chose to have in-depth conversations with a self-confessed shoe fetishist and a Buddhist monk who hadn't worn shoes in years. People who exhibit extreme behaviors usually do so because of an extreme need, belief or attitude. These factors also tend to exist in a given target group, albeit in much less pronounced ways that are much harder to detect. Thus talking to extreme users is another way to reveal latent needs within a target audience.

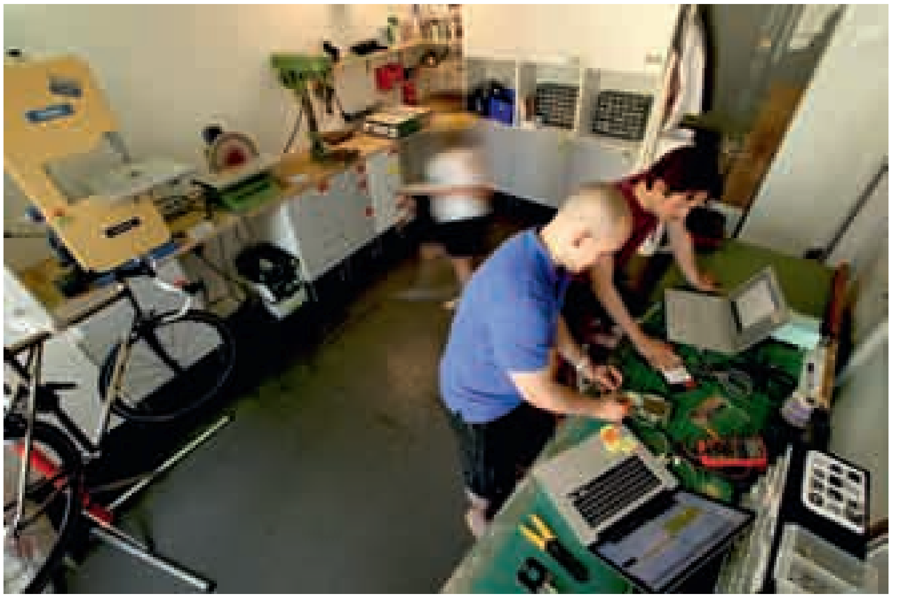
In the case of BBVA, the main limitation of customer interaction with ATMs wasn't the information, services, or benefits provided, but the way in which they were being delivered. Many customers felt the machines were insufficiently intuitive, transparent, or trustworthy. The design project, which set out to consider expanding the ATMs' feature set, quickly became about making the existing functions more accessible, more human and more intuitive. With BBVA, the design team developed a new machine with a much more human interaction. On this platform the bank can expand the benefits it delivers in the future.

See Beyond Face Value

As described, qualitative research can bring back fascinating insights from the field, yet they are not in a form that a typical organization can easily digest and act on. They usually take the form of anecdotes, photographs or individual quotes—not exactly data on which an enterprise is traditionally prepared to build a business strategy. What is crucial about inspirational research is how to interpret and use the results. In fact, translating user insights into business opportunities is the single most important—and at the same time underrated—capability in any innovation program.

After all, as a business looking for future solutions, we can ask the consumer a lot of questions, but we cannot expect them to serve up "The Answer" on the level that we need to base business decisions. The consumers in question are usually not experts in technology trends, nor are they up to speed on new ways to do business. Their horizon is naturally limited to their experience of the status quo and of having unconsciously adapted to it for better or for worse. Henry Ford's century old quote still rings true: "Had I asked people what they wanted, they would have said faster horses."

Instead of taking a consumer's input at face value, an innovation team must search out patterns, commonalities and differences among the usually hundreds of quotes and observations. This information is then aggregated to a point where it opens up interesting directions. None of these directions are guaranteed to lead to success, but as long as teams stay in short feedback loops with consumers, they will find out soon enough which direction resonates the strongest in a market.



Building prototypes enables teams to test and evaluate ideas early in the process

Look Beyond the Industry

Often, inspiration can be found in altogether different industries. This enables innovators to think outside of the proverbial box. In particular, looking outside of an industry helps if the best practices within a given industry have led to stagnation on the innovation front.

When some surgeons redesigned their emergency room operating procedures with IDEO's help, they found inspiration in how a Stock Car racing pit crew organizes its work at the racetrack. Among other things, the team learned that the pit crew actually works with redundant sets of tools that are sorted by likely usage scenarios (say, set one for a flat tire, set two for replacing parts of the car body, set three for suspension failure—all of them ostensibly containing an identical wrench among other tools). This saves critical time in gathering the right equipment when needed. Interestingly enough, the surgeons tested this same principle and found that it saved vital seconds in their operations as well. They implemented their own version of it.

Rapid and rough prototyping – sometimes it does not take much to bring a point across

3. BUILD OPERATING DEXTERITY

The authors of the IBM study essentially describe "operating dexterity" as the ability of an organization to adapt to change and its readiness to experiment with offer, structure, or business models. The CEOs responding to the study see this as a necessary organizational skill. "Operating dexterity" is thus shorthand for a whole range of organizational skills and mind-sets, some of which are crucial to bringing disruptive innovation to fruition.

Find Question Zero

An innovation initiative will only ever be as good as the question that it sets out to answer, or its brief. That is why it is crucial that a brief is shaped on the right level, meaning focused enough to yield the right outcomes, yet open enough to leave room for unexpected, disruptive results.

When shaping briefs, we often find that clients tend to roll part of an expected solution into the brief, which unnecessarily constrains it and leaves a team working on answering it with only one direction in which to move. In our experience, the best way to create a rich brief is to jointly determine what "Question Zero" is in the context of the challenge. By this we mean the underlying human question, the one that asks for the need in the market on the level of an individual within the target audience.

For instance, when working for a publishing house specializing in textbooks for university students, the IDEO project team discussed many possible starting points. Briefs such as "How do we sell more textbooks?" miss the point, because it would assume that both the media (books) and the business model (selling) would remain untouched, a seriously limiting assumption in times of fluid media and drastically evolving business models.

The brief we eventually agreed upon was "How do we better support students in their learning?" It enabled the team to safely move within the boundaries of the purpose of our client's organization, yet left room to explore new media, new ways to communicate with students and to facilitate communication among them as well.

Having established how students learn and what their explicit and latent needs were around this—often social—activity, the team was then able to delve into how to best support students, building on not one technology, but a number of media channels in which the students were already engaged. This meant a new way of doing business for our client, but little or no learning curve for users and thus the assurance of speedy adoption.

Failure is Essential For Iteration

After bringing back interesting insights from the field, with qualitative data to back them up, companies can establish their value by prototyping and testing ideas quickly. Early prototyping—albeit with crude tools—that allows for refining potential solutions step by step, is essential to evaluating how technically feasible, financially viable, and ultimately desirable an offering might be.

Prototyping as an innovation tool is not limited to products, but can include spaces, services and processes as well. It is important that prototyping should take place early and that innovators should only invest as much time as needed to answer the most important question at that moment, but no more.

The surgeons mentioned above learned from unlikely sources in their quest for innovation opportunities. What they did next was to prototype and iterate their way to an innovative solution. They built and tested numerous ideas that failed, as was expected, before finding the ideal solution.

Often seen as negative, failure wasn't a problem in this case. This is because they prototyped many ideas, so they had a number from which to choose and secondly, because failure occurred early on in the process. They immediately started prototyping potential solutions and testing them out in a simulated operating room. If an idea didn't work, they found out within days or weeks, rather than months or years. This meant little time was lost in testing, little money spent to that point and little opportunity to fall in love with an unrealistic idea.

Demystify Failure

Of course, no one wants to fail spectacularly after having spent millions in development costs. But when failure happens early, before major capital has been invested, and leads to more insights, a company has a far greater chance of success in the next iteration in the marketplace.

Too many traditional organizations still consider failure to be taboo—and thus avoid failing at all costs, including innovation. The primary problem seems to be that initiatives that fail today tend to fail too late, which makes their failure expensive. Because of that, all failure is perceived as bad and damaging to images and careers. Yet the right kind of failure should be a requirement. As Seth Godin wrote in *The Purple Cow*, "Most learning, especially most organizational learning, occurs through trial and error. Error occurs whether you want it to or not. Error is difficult to avoid. It's not clear that research or preparation have an enormous impact on error, especially marketing error. Error is clearly not in short supply. Trial, on the other hand, is quite scarce, especially in some organizations. People mistakenly believe that one way to successfully avoid error is to avoid trial. We need more trial."

Start Small and Scale Up

Having developed an innovative concept deemed good enough to bring to market, there usually is a lot of learning still left to do on the path to perfecting it. In order to learn in the real world while mitigating the risk of failure, many companies have perfected the art of starting on a small scale and then scaling up as they go along.

3M decided to launch the Post-It note as a new product in 1977, but it failed because consumers had not tried the product. Convinced that the overall concept was sound, a year later 3M distributed free samples to residents of Boise, Idaho in the United States. Nine out of ten people who tried the notes said that they would buy the product. By 1980, Post-Its were being sold nationwide in the US The rest is history.

3M managed to learn in a controlled setting about multiple dimensions of their new offering: They tested how to best advertise it, who to speak to and where to place their products. They learned about how the product is actually used and how it is being received and how they could improve it—all of this by focusing on one small town in Idaho. The "Boise Blitz", as it became reputedly known within 3M, was a very intelligent step taken to bridge the gap between the R&D lab and a nationwide rollout, and a great example of starting small and scaling up.

Making an organization as innovative as 3M manages to be might seem a long way out for most business leaders. To make innovation a mindset for all staff is not an easy task by any stretch of the imagination. On the positive side though, usually no major structural changes are required: no re-org, no large-scale hiring efforts or changed management processes necessarily stand in the way of a company realizing its innovation potential. All that is needed is for an inspired business leader to put innovation on everyone's agenda, to lead by example and to give their employees the space and permission to be creative in a safe environment. That is a hard enough job as it is.

The journey is indeed rather easy to start for any leader. True to the spirit of design thinking, any company's approach to innovation should be imagined, prototyped, iterated and evaluated on a small scale first in order to evolve into a bespoke, appropriately adapted organizational skill before being scaled up. In fact, a few easy experiments outlined in the side box can serve as first steps. See them as suggestions on how to get the first few miles on the way to New Orleans.

Starting the Innovation Journey

Many organizations wonder where or how to start becoming more innovative, user-centered, and adaptable.

Here are five experiments to try:

1. Tap the everyday experts within your organization.

You already know more than you think you do. Every organization has employees who are in daily contact with customers, clients and partners. They may be sales staff, maintenance technicians, or call center employees. These are the people to whom the rest of the organization should listen closely and gather insights from.

2. Get to know a few customers yourself, one by one.
Join your market research teams on home visits to your customers. Get to know a few of them and start seeing patterns emerge that point to new business opportunities. It's rewarding, revealing, and sets an example within your organization.

3. Try out a 'no numbers' rule. Run an innovation initiative based purely on *qualitative* metrics. Learn to appreciate user anecdotes, to trust expert intuition, and to channel your staff's passion. Think people instead of pie charts. Chances are the initiative will move faster, cover more ground, and unearth opportunities earlier than initiatives burdened with the need to justify their efforts quantitatively every step along the way.

4. Ask innovation teams to question everything. Within the purpose of the brand, your teams should be free to question every aspect of the status quo. Acknowledge that your company is not defined through the current product or service, nor through a business model. They are both just tools that help you achieve your purpose. Don't let them study the competition, but instead let them figure out what it would take to disrupt your market. Results might be far out, but they will serve well to set a direction for you to head in.

5. Let your teams experiment early and safely. Request that any innovation initiative include some sort of prototype or experiment that shows its progress at every check-in meeting. Make it clear that you would like to see all results, including those that failed during testing, in order to talk about what the team learned. Challenge employees to discuss unfinished work with customers as soon as possible and to make potential results tangible, even if they feel half-baked. Allow for "safe" failures and celebrate eventual successes.

BBVA

Innovative Culture: Values, Principles and Practices of Senior Executives in Highly Innovative Companies

Joaquim Vilá
IESE Business School

1. INTRODUCTION

One of the reasons why many companies and institutions are so cautious about increasing their innovative capacity is that innovation represents a seemingly complex problem for the senior executives who must initiate and direct the changes that robust innovation requires. Innovating entails transforming new ideas into tangible results. There are three important aspects to be considered in this approach to innovation. First of all, merely coming up with new ideas is not innovation; if the ideas are not implemented, the desired effect cannot be achieved. Churning out new ideas for the sake of it does not make us more innovative. Secondly, when a new idea can be applied without changing a company's day-to-day reality, it falls under the heading of continuous improvement rather than innovation. Continuous improvement and innovation both make a contribution, a market impact or an internal improvement, but the difference between innovation and improvement is that the implementation of an innovative idea requires a change of tack, a significant modification of the company's former routine. Thirdly, and in connection with my second point, innovation demands and relies on new developments: a new contribution to basic knowledge, a new process or a change of direction, some kind of internal development that complements the original new idea. Innovating today makes it easier to innovate tomorrow.

The attention of many top executives is largely focused on the urgent matters of everyday management. Innovation is often put on the back burner for, although it is considered an important issue, it does not have the same priority as other more pressing concerns. If company directors were aware of the fact that postponing innovation today would undermine their ability to innovate tomorrow, innovation would (or should) be given higher priority and create a predisposition among senior executives to lay the foundations that will make it possible.

When a senior management team becomes aware of the transformations that innovation requires, it identifies a series of challenges that must be faced. These reflections usually address issues such as:

- The need to protect innovation from potential threats that would prevent the company from taking off in a new direction: management systems (like budget planning), chains of command opposed to change, organisational inertia and other determining factors carried over from the previous system.

- In order to be credible, innovation must produce consistent results. It cannot be a fortuitous, one-time contribution. Innovation must become a permanent and deeply embedded fixture of organisational culture.
- Coming up with new ideas requires a certain attitude (thinking big, challenging things, seeing more possibilities than problems, etc.) which in many cases is the exact opposite of how people have been taught to think in the corporate context. How can we make innovation both credible and appealing to the people we work with?
- Creativity will only flourish if there is a change in management style; executives have to realise that the rules which govern everyday management activities—such as exercising bureaucratic control, communicating via the established chain of command, equating the value of a person's idea with his/her status in the corporate hierarchy, and other similar rules that allowed operations to grow steadily and predictably in the past—cannot be applied when innovation is on the agenda.

Faced with such a scenario, most executives tend to pull back, look away and hope that the profit margin can be increased by discovering a new way to cut costs. The rest, those executives who are willing to face the challenges involved in developing new capacities in the company, take comfort in the knowledge that innovation has prospered in traditional companies in almost every sector. The solution to the puzzle is based on principles which, while not purely intuitive, do have great internal logic and are also within reach of any management team that is willing to change, even if it means starting with themselves.

There are two key aspects to any organisational solution which aims to achieve robust innovation in a conventional company. First of all, innovation is a challenge that has much more to do with human resources management than with technical or design problems. Developing the capacity to innovate requires a paradigm shift in managerial strategy and in how people act. Secondly, and more importantly, innovation targets the pinnacle of the management pyramid; it is a matter of transforming leadership that brooks no delegation. Moving towards a completely innovative corporate culture calls for new managerial skills and practices which, up until now, have not been considered vital qualifications for top corporate management positions. However, an examination of the working methods of senior executives at innovative companies reveals their awareness of one important fact: that the singularity of their companies begins with a change in themselves.

2. INNOVATIVE CULTURE

When a company first launches an innovation process, it is usually focused on the key performance areas within its business strategy, such as product, service aspects or other business development issues. In every case, innovation is successfully brought to market by combining efforts in several different areas. The key is to spread the responsibility for innovation as much as possible rather than limiting it to a small group of individuals or specialised departments.

As they acquire more experience with innovation, company directors begin to see that any area of the business can be revitalised by applying new ideas, and that innovation can be used to work towards a number of goals, not just competitiveness. At this point, innovation ceases to be a specific responsibility assigned to one team or department. The vision of immersion in innovation becomes a project to make innovation a broad core value for the entire company. The approach becomes one of wide-scale innovation, which yields regular results. This represents a major change of perspective. The pursuit of management innovation goes far beyond the initial intention of managing innovation within a limited area. The management soon sees the advantages of creating an environment where everyone in the company can contribute, and creating an innovative corporate culture becomes one of the top priorities on the agendas of the senior executives and the board of directors.

The most widespread notion of organisational culture is that it is the sum of values and beliefs which people have acquired over time and which dictate the appropriate standards of behaviour within a company. Edgar Schein, an MIT professor famed for his valuable contributions in this field, defined culture as a "pattern of basic assumptions that a given group has invented, discovered or developed in learning to cope with its problems of external adaptation and internal integration and that have worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think and feel in relation to those problems".

How can management implement and maintain an innovative culture? If the directors think that culture can only be changed by directly altering the values and beliefs that employees have learned and share, the task can seem dauntingly complex. Values and beliefs are a part of people; they have become a part of each person, ingrained in his/her mind, and are therefore hard to change if a direct approach is taken. Yet senior executives, in their daily routine and by their example—in other words, the way they shoulder their responsibilities and interact, formally or informally, with the people around them—continually show what they value in, and expect from, their co-workers. The values and beliefs that comprise corporate culture are largely the result of accumulated experiences over time deriving from a specific management style.[1] This suggests the existence of a much easier way to bring about change. The management can promote a culture favourable to innovation by changing the way they lead and interact with their employees. This basically means that organisational progress (the way leadership responsibilities are defined, different management systems, and the manner of integrating and coordinating management practices, etc.) is achieved by recognising people's contribution to creativity and innovation. Clayton Christensen (1999), a professor at Harvard Business School and a leading expert on innovation, reached a similar conclusion when he stated that "culture is comprised of processes, or ways of working together, and of shared criteria for decision-making [...] which have been employed so successfully so often, that they come to be adopted by assumption".

This highlights one of the most serious shortcomings of many companies—the ability to introduce organisational changes that facilitate the attainment of concrete objectives while also opening up new possibilities. Implementing management systems specifically geared towards

[1] The idea of culture as a result of the sedimentation of management practices over time was proposed by Esteban Masifern, lecturer at the IESE Business School, in the late 1980s. The conversations with Carlos Cavallé, also an IESE lecturer, in 2007 were instrumental in determining the importance of the behavioural attributes of senior executives beyond their mere personality traits.

encouraging innovation is a much safer and more solid option than attempting to change the way people think. The best way to create an innovation-oriented culture is not to replace executives opposed to change or wait for a new generation of executives to step up whose talents and values are more in keeping with the new strategy. Companies need to accept that the most effective way of creating the culture they want is to change their management practices, with a clear orientation towards the desired goal.

The best tactic for creating an innovative culture is to introduce management practices (progressively spreading to employees at every level through various channels) that promote the desired innovative, entrepreneurial and creative behaviour. By changing our management style (in each and every aspect: how we set goals, plan, allocate resources, evaluate co-workers, reward and recognise, assign degrees of responsibility and autonomy, manage information, etc.), we are shaping corporate culture, whether consciously or, as is often the case, unconsciously. Once again, these activities target the highest echelons of our organisations.

The aim of this chapter is to outline the best practices of highly innovative companies which any business corporation can look to as examples to follow. I will focus on aspects that point directly to various facets of the role of senior management and leadership styles, because I am convinced that this is the driving force behind the innovative capacity of many companies. Obviously, these case studies will not translate directly into specific implications for innovation management in every company, but the lessons that can be learned from their example will shed light on how senior executives can make their interaction with employees more consistent and predictable and so lay the foundations for an innovative culture.[2]

When examining the best practices of highly innovative companies, the executives' first reaction is to point out the unique features of their specific company. "Yes, that's all well and good, but our company is different." Senior executives tend to doubt the feasibility of implementing the management practices of highly innovative companies in their own organisations. Although it is true that some of these practices are only applicable at advanced levels of innovative culture, the general rules for encouraging people to move towards innovation can be applied in almost any field. After analysing the management levers used in innovative companies and comparing them to those of less innovative organisations, Tushman and O'Reilly (2002) noticed that all innovative companies shared a similar set of management principles, regardless of their geographic location or field of activity. The conclusion is that, although the implementation of a new management style should be tailored to the specific needs of each company, the basic principles which govern the pursuit of innovation are fairly universal.

Another issue that management teams must face is determining the most appropriate approach. The innovative culture of companies which excel at innovation is like a hologram. One can take a variety of different approaches to the phenomenon and still get the feeling that they all effectively point to the same reality. It is relatively easy to tell if a company is innovative by observing different aspects of its dominant culture. The implication is that, although the companies described here may have used different approaches and management strategies to promote innovation, the underlying patterns

[2] In another publication, I have described and discussed different management approaches for moving towards an innovative culture (Vilá, 2008).

of managerial logic and criteria have a great deal in common. I believe that the solidity of the innovative nature of these companies derives from the coherence of the rules of the game that shape their innovative cultures (values, principles and practices). As I have already mentioned, this essay does not aim to discuss the different approaches and mechanisms which allow a company to move towards an innovative culture; rather, my intention is to focus on the best practices of senior executives in highly innovative companies.

3. CONDUCT OF SENIOR EXECUTIVES IN THE AREA OF INNOVATION

In the course of their interactions with other employees, senior executives give clear indications of their personal preferences. The way in which a management team interacts with the rest of the organisation shapes the values and expectations of its workers with regard to their conduct and attitude towards innovation. This performance is strictly analysed in the light of the senior management's context of activity. The effect on culture is greater in situations where employees understand that an innovative approach can be more beneficial (for example, in seeking solutions to important problems, defining business development strategies, evaluations when allocating strategic funds, responding to unforeseen opportunities, coming up with a new strategic approach when the market undergoes major changes, etc.). In cases where the context calls for a management strategy governed more by the principles of everyday operations, employees understand that it is not their place to interfere in executive preferences in the area of innovation.

The leadership style and conduct of senior executives is a vital factor for stimulating innovative capacity (Deschamps, 2008).[3] The most typical traits of a senior executive's favourable personal behaviour (as an individual) towards innovation can be analysed from several different angles. The approach I have chosen here is to analyse what drives their efforts to spearhead the transformation of their companies, how they direct the progress towards that goal, and how they face the problems that crop up during the evolution towards innovation. In this text I will not address the more collective aspects, which are also essential for creating and maintaining an innovative culture but which require a more detailed and separate analysis. Collective aspects are those which require the coordinated efforts of the management team, such as the creation of a favourable environment (creating an organisation for innovating that is compatible with the organisation of everyday activities) or the design and implementation of a senior management process that generates business and management innovation initiatives with a broad scope and potential implications for any area of management. I will now proceed to describe the three fundamental pillars of the actions of top-ranking executives, the cornerstone of the process for inculcating the values, principles and practices that will determine a company's capacity for innovation and quality.

They are driven by the pursuit of a goal, ideal or dream that makes sense to those who have to make it possible

Senior executives act as leaders who seek to make an impact with highly ambitious initiatives that will offer unique contributions. On the same day Steve Jobs announced the details of the iPad, his Apple co-founder Steve Wozniak addressed a smaller, less

[3] Jean-Philippe Deschamps (2008) was a valuable source of inspiration in my efforts to pinpoint the different ways in which senior executives contribute to innovation.

formal group at the Laxson Auditorium in Chino, California. The brilliant engineer spoke about the late 1970s and early 80s, when he embodied a new way of thinking, far removed from the world of mainframes and minicomputers, when he and Jobs helped birth the personal-computer industry. "I was turned on that little guys were going to do something of more value than the big corporations," recalls Wozniak. "My friend Steve Jobs... was always interested in doing things that would change the world. He was a move-the-world-forward kind of guy." Source: http://www.successmagazine.com/steve-jobs-master-of-innovation/PARAMS/article/1054/channel/22

Senior executives are strongly motivated to overcome challenges, and their efforts also make perfect sense to their employees. A challenge is a tangible goal that focuses everyone's attention on one point, and in many cases it is a more effective stimulus for innovation than extrinsic incentives such as performance-based salary packages. (See text box 1 on Ratan Tata.) A remarkable case in point is Akio Morita and Masaru Ibuka, the founders of Sony, who expressed their unease and concern about not having a big enough goal for their product development engineers. In 1952, Ibuka wondered, "Would tape recorders be challenge enough for them, motivate them to use their best abilities, or let them grow to their full potentials?" This ambitious diligence paved the way for ground-breaking product launches like the Trinitron television, the Walkman and many others.

Senior executives at the helm of innovative companies are capable of rallying the efforts of the entire organisation around ideals that will have social benefits and also motivate people to make their own contributions. Franck Riboud, chairman and CEO of Danone since 1992, sums up his company's great ambition as the desire to "bring health through food to as many people as possible". These executives are very passionate about the mission that gives meaning to their company and to innovation. (See text box 2 on John Mackey of Whole Foods Market.) Their clarity of purpose and energy create an infectious enthusiasm. The goal they pursue is a powerful internal force that propels them and the company forward. The initial boost is fundamental. They are attracted by the chance to do something unique. The analysis

Text Box 1. The Urge to Pursue a Dream

Ratan Tata started with a Dream. His dream was to provide every Indian an affordable and safe means of family transportation. What would be affordable for a middle class Indian family? At that point in time (2007), Maruti Udyog was the cheapest car at USD 7,000. Ratan Tata felt that a car of up to 1 lac rupees (100,000 rupees, roughly USD 2,500) would be affordable. So he set this goal to his engineers at the Tata Motors. "It's a dream project for the man who has an impeccable record of accomplishing things which he dreams and the name of the car is Nano," claimed Mr. Sunil Sinha, CEO of the Tata Quality Management Services.

It was never meant to be a 1 lac rupee ($2,500) car; that happened by circumstance. Ratan Tata was interviewed by *Financial Times* at the Geneva Motor Show: "I talked about this future product as a low-cost car. I was asked how much it would cost and I said about 1 lac rupees. The next day the *Financial Times* had a headline to the effect that the Tatas are to produce an Rs100,000 car (USD 2,500). My immediate reaction was to issue a rebuttal, to clarify that that was not exactly what I had said. Then I thought, I did say it would be around that figure, so why don't we just take that as a target. When I came back our people were aghast, but we had our goal." Ratan Tata had a choice to call the newspaper and correct the mistake. Instead he took it as a challenge for himself and his engineers.

Source: "The Making of the Nano", Christabelle Noronha. January 2008. Tata Leadership with Trust. • Contributed by Preeti Sharma, Gemba-2010. http://www.tata.com/media/interviews/inside.aspx?artid=Sd75BUBmzSM.

Text Box 2. A Mission to Channel Innovation

John Mackey, CEO of Whole Foods Market, co-founder of the company in 1978, has pushed the company to exceptional growth and success (CAGR of 17% in sales and of 14% in profits in the period 2001-2010). In a conversation he had in April 2010 at Darden School of Business with Professor R. Edward Freeman, he reflected upon a renewed sense of purpose for himself and Whole Foods Market: "After a great deal of soul searching into some of my deepest passions about the purpose and meaning of my own life, I came to the conclusion about 18 months ago that continuing to help lead and evolve Whole Foods Market is exactly what I most want to do, and I recommitted to the company for the long term.

I have become especially excited about the potential to help improve the lives of millions of people through better education about the principles of really healthy eating and I'm helping lead these rapidly evolving efforts at Whole Foods Market. We have many exciting healthy eating and wellness initiatives in development that I believe are really going to help people live healthier and more vital lives, and I look forward to sharing them with everyone during the next few years."

Source: The CEO's Blog, "Darden School of Business Conversation" by John Mackey, 26 July 2010• Contributed by Dana Page, Gemba-2010. http://www2.wholefoodsmarket.com/blogs/jmackey/

of constraints, though important, takes a back seat and always comes afterwards.

At times, the goal that inspires the collective efforts of the entire workforce is not a dream or a mission with a definite social contribution in mind, but an ambitious strategy which senior executives take care to renew and reinvent over time. This is often the case of consolidated companies with a long history of innovation, as illustrated by examples based on 3M, General Electric or Procter & Gamble. The key ingredient in every case, however, is that the goal must be shared in order to inspire a team effort, and for that to happen it must be "tangible" or sufficiently concrete, and it must also be appealing to those who are going to make it possible.

How they direct the progress towards the goal

The management style of an executive team is a fundamental factor for understanding a company's capacity for innovation. Senior executives in highly innovative companies share values that encourage experimentation and learning. The history of Apple, one of today's most innovative companies, is dotted with failed attempts, such as the pioneering Newton PDA or (though not all would agree) Apple TV and Apple Pippin. However, Apple learned valuable lessons from its mistakes and applied them to later developments. Thus, the mobile telephone developed with Motorola, the E790 iTunes, and the Lisa computer could be described as unsuccessful projects, but they made different contributions that were successfully used in the revolutionary iPhone. Jesús Vega, former director of human resources at Zara-Inditex, once said, "The important thing is not to avoid the pitfall of error but to climb out of the hole once we have fallen in."

Senior executives are firmly committed to exploring new ways of operating, even in the absence of pressing problems, and are open to unconventional solutions and ways of thinking. (See text box 3 on James Dyson, founder of Dyson.) IKEA's management encourages employees to search for new and improved ways of doing things in every aspect of their jobs. Even though IKEA is the undisputed leader in the global furniture market, it still promotes continuous improvement and change. The people working for the IKEA group often feel just as motivated by the mission to find new ways of achieving goals as they are by the goals themselves. They are inspired by discovery and constantly looking forward to the next new challenge. In a different context, Guy Laliberté, founder and CEO of the Cirque du Soleil, refuses to rest on the laurels of past success: "A typical day at the office for me begins by asking: What is impossible that I'm going to do today?"

This positive predisposition to explore even the risky options must be accompanied by a fundamental ingredient of any innovative culture: the acceptance of a degree of uncertainty and, consequently, a tolerance for good-faith errors. The words of William L. McKnight, CEO of 3M from 1929 to 1949 and chairman of the board of directors from 1949 to 1996, reveal his superior management style: "Mistakes will be made. But if a person is essentially right, the mistakes he or she makes are not as serious, in the long run, as the mistakes management will make if it is dictatorial and undertakes to tell those under its authority exactly how they must do their job. Management that is destructively critical when mistakes are made kills initiative, and it is essential that we have many people with initiative if we are to continue to grow."

Highly innovative companies accept the fact that relevant information, good ideas and initiatives can come from sources outside the company. Keeping an open mind to the outside world can become a part of formal corporate procedure, as in the case of Procter & Gamble's "Connect + Develop" strategy or IDEO's "Deep Dive" brainstorming technique; it can inspire an approach based on strategic business partnerships, such as that embraced by Amazon, Nike or Federal Express; or it may take the form of employee initiatives supported by the management, like the participation in social networks and forums practised by the employees of Sun Microsystems, or an open invitation for external agents—customers, for example—to suggest ideas, a strategy used by many companies. Senior executives urge their employees not only to look outside the company but also to become keen and active observers and learn to identify business opportunities in the tangle of weak signals bouncing around the market. This is exactly what goes on in the business opportunities evaluation process of the "Imagination Breakthrough" initiative led by the CEO of General Electric, Jeffrey Immelt, which aims to generate ideas that will boost the company's growth.

A flexible approach is vital for progressing towards the ambitious, challenging goal. Management at innovative companies shows an open mind and is willing to work proactively and humbly to find other

Text Box 3. Urging Employees Not to Settle for the First Easy Solution

James Dyson, founder of Dyson Ltd., UK, built on a vacuum cleaner that uses centrifugal force to separate dirt from air he had invented in 1979. Having gradually stepped away from the day-to-day running of the business, he's spent the past decade trying to create the perfect environment for innovation. He's determined not to be an owner marooned from the workforce. "One of the most important things is I spend time, not in my glass office in Wiltshire, but going among creative people, not just the engineers, making sure they're doing creative things. I don't mean I go around like a policeman, but more just encouraging creativity."

Offering praise, encouraging workers to take the challenging route rather than the obvious one, and taking an active role himself are pillars Dyson has built the company on. "What I'm talking about is people having the courage to take risks, make a mistake and be ridiculed. They should have no fear of doing something that's not 'normal' or 'sensible' and worry I'll clip them round the ear and say 'don't be so bloody stupid'," he claims.

Source: "James Dyson", by Ian Wallis. *Growing Business Online*. 1/4/2004. • Contributed by Federico Ciardelli, Gemba-2009. http://www.growingbusiness.co.uk/06959143454587923447/james-dyson.html alternatives for moving forward. The managers of Starbucks tell their employees, "When presented with negative feedback from a customer, recognise that you may have an opportunity to actually strengthen that relationship." When employees see that management takes a healthy interest in all feedback, both positive and negative, appreciating its value becomes easier for the workers themselves.

Evaluation for the purpose of learning fosters innovation and enhances the ability to achieve the established goal (Senge, 1998). At Pixar Animation, executives like Brad Bird, who can boast a string of blockbusters such as *The Incredibles*, *Cars* and *Ratatouille,* treat their employees as equals. At team meetings, everyone receives positive or negative feedback on their work before the entire group. Everyone is encouraged to participate and express their opinions about the work of others. Brad Bird's introduction of this practice at Pixar has raised the bar in terms of the quality of the company's output. (See text box 4 on Brad Bird, director at Pixar.)

These values, principles and practices used by senior executives in highly innovative companies are not limited to a specific industry or geographic region. They can be found in large corporations with a technological background (like GE under the leadership of Jeffrey Immelt) and companies whose markets are constantly evolving (such as Inditex, owner of the Zara chain and a leader in the global fashion industry, in which the founder Amancio Ortega has played a pivotal role), but also in traditional businesses like the Mexican cement company CEMEX and even in much smaller enterprises. At Metalquimia, a second-generation family business that manufactures meat-processing equipment, the strong personal commitment of executive director Josep Lagares has been the driving force behind initiatives to strengthen the company's capacity for innovation. He heads up the product innovation and process committees at Metalquimia, the entire workforce has received training in creative techniques, and he has introduced a computer application for managing creativity and innovation that encourages everyone to contribute—"just another step towards my goal of socialising creativity". In April 2010, Metalquimia introduced Quick-Dry-Slice, a revolutionary technology for speeding up the drying and curing process of sliced meat products which will reinforce the company's already solid position in the global market.[4] The executive director is the primary catalyst for the process of democratising innovation within this company.

Text Box 4. Value Based on Validity, Not Hierarchic Status

Brad Bird, director at Pixar: "From the beginning, I pushed and analyzed each person's work in front of everybody. And they didn't speak up. One day, I did my thing, and one of the guys sighed. I shouted, 'What was that?' And he said, 'Nothing man, it's OK.' And I said, 'No, you sighed. Clearly, you disagree with something I did there. Show me what you're thinking. I might not have it right. You might. Show us.' So he came up, and I handed him the dry-erase marker. He erased what I did. Then he did something different and explained why he thought it ought to be that way. I said, 'That's better than what I did. Great.' Everybody saw that he didn't get his head chopped off. And our learning curve went straight up. By the end of the film, that animation team was much stronger than at the beginning, because we had all learned from each other's strengths. But it took two months for people to feel safe enough to speak up."

Source: "Innovation lessons from Pixar: An interview with Oscar-winning director Brad Bird", *McKinsey Quarterly*, April 2008 • Contributed by Edinardo Figueiredo, Gemba-2010. http://www.mckinseyquarterly.com/innovation_lessons_from_pixar_an_interview_with_oscar-winning_director_brad_bird_2127

[4] http://www.metalquimia.com/productes.php?idm=3&subpagina=20

How they deal with adversity

As with any transformation of the status quo, innovation entails dealing with unforeseen events. The reactions of senior executives to difficulties that the company must overcome as it pursues its goal set an example for the other employees by revealing the values and preferences of the management, and their response therefore has a direct effect on the company's innovative culture.

Senior executives are persistent and show unwavering support for an initiative even when the early results are disheartening. In the early years of Sony's history, Akio Morita and Masaru Ibuka were warned by their accountants that they were investing too heavily in new developments and even putting the company's financial viability at risk. The executives' response was: "Just be patient a little longer and we will make a fortune." Another example of perseverance is Carlos Sumarroca, who refused to listen to a scientific/technical team's repeated advice to scrap his project; today, Agromillora Catalana is one of the world's leading tree and rootstock suppliers.

In highly innovative companies, the phrase "it's impossible" does not exist. When faced with recurring failures, they simply say, "We don't know how to do it right." Some of the solutions that are now part of the IKEA business model came about thanks to dogged attempts to solve problems that could have sunk the company. Their obstinate insistence on keeping costs low and finding an original solution to a transport strike led them to the ground-breaking idea of having customers transport and assemble the furniture in their own homes.

Innovation requires persistence. "If at first you don't succeed, try, try and try again." The IKEA motto, "Never say never encapsulates the company's positive obstinacy, perseverance and determination to attain goals and never give up. However, this persistence is tempered by limits and principles that make it balanced and financially viable for the business (Ingvar Kamprad, founder of IKEA, in his "Testament of a Furniture Dealer"). This willpower is considered more effective at the team level, where there must be mutual trust and general consensus before coordinated, decisive action can be taken.

At Amazon, the management encourages employees working in areas of innovation to pursue ideas and concepts even if they do not yield short-term results. "You need to be willing to be misunderstood for a long period of time. Innovation is hard for large companies because you need to be long-term oriented." Commenting on the need to think long-term, Jeff Bezos, founder and CEO of Amazon, stated with conviction, "If you take a short-term approach, then you are constantly stuck with trying to deal with minutiae. I hate to see someone 'sticking to their knitting' because they're incapable of taking chances."

In highly innovative companies, senior executives are visibly self-confident, but they also give clear indications of having confidence in the work of entrepreneurs. At 3M, an entrepreneurial spirit is encouraged by the application of a fundamental principle: freedom to pursue ground-breaking ideas in the workplace. Researchers at 3M can spend up to 15% of their time working on projects of their choice, with no interference from management. Thanks to the atmosphere of freedom created by this policy, a 3M scientist named Art Fry was able to develop the famous Post-It note in 1973.

Executives in charge of innovative companies are willing to step outside their comfort zone to make innovation possible.

Text Box 5. Senior Executives Are Willing to Change Even Themselves

"Innovation should be at the basis of everything we do, not only the product formulation, but everything that touches the consumer. It's the concept of innovating in all the ways that touch the consumer life. P&G has moved from a definition of innovation, which was probably a bit narrow, focusing on the product, to a concept of innovation that is a lot wider.

Innovation is becoming more demanding and pace is becoming shorter. We will never have enough resources, enough ideas internally to deal with it so quickly and broadly. P&G has set a great goal: to get about half of the ideas from outside. When you've made that decision, you need to change the way you work and make those connections. People have to want to come to you with an idea, they have to believe that you'll listen and give them a chance. They have to believe that you're prepared to change some of your processes to accommodate new ideas."

Gianni Ciserani, Vice President and General Manager of Procter & Gamble UK and Ireland

Source: Brand Strategy, Published: 01 November 2005 "Learning to Listen to New Ideas: Interview with Gianni Ciserani" • Contributed by Andrey Lankovich, Gemba-2008 and Manuel Lapeira Gemba-M-2010. http://www.mad.co.uk/Main/News/Disciplines/Marketing/Articles/5f42852a91a6407db37cf0114d285395/Gianni-Ciserani-the-Italian-job.html

Their confidence in the benefits of innovation, and their commitment to making it possible, are evidenced by the fact that senior executives are willing to consider initiatives which entail changes that may affect them personally, altering their own quality of life in the company. They do not hide behind the shield of hierarchical authority. They are just another cog in the machine, subject to the same rules as everyone else. When Sony decided that it needed to study the unique traits of the world's most advanced consumer electronics market firsthand, Akio Morita moved to New York with his entire family in July 1963—an eloquent example of personal commitment.

When the leaders of highly innovative companies are the most tangible examples of personal commitment, it creates the necessary climate for bringing about a change in employee values and beliefs that will favour innovation. Senior executives face adversity with courage and bravery, a clear reflection of their high self-confidence. Their commitment leads them to play an active role in innovation governance, supervision and management, and their intense involvement and discipline in spearheading the process is understood by everyone as proof of their serious sense of purpose. Logically, the tangible commitment of corporate leaders ends up becoming a necessary ingredient in both the transition towards and the firm consolidation of an innovative culture.

Jeffrey Immelt, GE's CEO, fosters GE's commitment to innovation by leading through example. As the world is resetting itself, Immelt says that his role is to turn fear into confidence.[5] At a speech at Dartmouth College he offered a grounded playbook for leadership in the tough present. "Good leaders," he said, "are very curious, and they spend a lot of time trying to learn things." Immelt tries to set aside 20% of his time for thinking and re-conceptualising. He also said that "every innovation leader at GE will meet with me at least once, and a great part of my job is that I get to go places to pick up that next best idea". Yet, like his predecessor Jack Welch, known for deep dives into operating details, Immelt said he is involved with 40 projects at GE that represent "imagination breakthroughs". He has developed his own guidelines to judge new opportunities and progress.[6] Good leaders manage by setting boundaries with freedom in the middle. "The boundaries are commitment, passion, trust, and teamwork. Within those guidelines, there's plenty of freedom. But no one can cross those four boundaries."

[5] http://ge.ecomagination.com/ and http://www.fastcompany.com/blog/lewis-perkins/semantics-sustainability/sustainable-strategy-goes-mainstream

[6] http://www.fastcompany.com/magazine/84/pr.html

Jeffrey Immelt, in a conversation with *Fast Company* editor-in-chief John A. Byrne in the CEO's conference room at GE headquarters in Fairfield, Connecticut, was asked: "What have you learned so far in the job as GE's CEO?" Immelt responded: "One of the things Jack Welch said early on that I think is totally right is: It's a marathon. It's not a sprint. All these books about the first 90 days are kind of rubbish in many ways. You have to have a plan. You have to stick with it. You have to modify it at times, but every day you've got to get up and play hard. Jack used to see me running around, even after he left, and he'd say to me, 'Remember, it's a marathon. Ten years. Fifteen years. You've got to get up every day with a new idea, a new spin, and you've got to bring it in here every day.' I always kind of knew that, but until you're right in the middle of it, you never get it. His advice was right. It's the sustained ability to change that really counts."

4. IN CONCLUSION

Innovation requires making changes in the way a company works so that new ideas can be transformed into results. These changes must be understood and supported by the majority of the company's employees. Moreover, innovation today is not confined to the technological sector or to new product launches, and building a solid support base for innovation is therefore a priority for senior executives. However, modifying values and beliefs to create a favourable environment for innovation seems like an arduous task at first glance.

The organisational culture of highly innovative companies suggests that the most effective tool for changing mindsets is management style, and the behaviour of the company's top executives is a crucial factor. How senior executives conduct themselves in situations involving innovation, whether consciously or unconsciously, has a direct effect on the employees' perception of the values and beliefs they are expected to embrace, and so their actions determine the degree to which the corporate culture is more or less favourable for innovation. The task of mobilising employees to pursue innovation begins with a change in attitude and conduct at the highest levels of authority in our organisations.

A study based on the role of senior executives in highly innovative companies suggests that their conduct can influence the creation of two important but very different kinds of innovative cultures (Vilà, 2008). The first and most obvious is an innovative culture that is largely defined by the personality of a founder (or a small team of founders) with a strong entrepreneurial spirit. The values and beliefs of companies like Apple or IKEA are closely bound up with the actions and traits of their respective managements. It would be impossible to conceive Apple's early innovation without Steve Jobs, or IKEA without Ingvar Kamprad, or Sony without Akio Morita and Masaru Ibuka. The second kind is a culture that transcends the figure of the founder and endures over time. We might say that in this case the values have become part of the company's DNA; they have been institutionalised. Does anyone know the names of the founders of companies like 3M or Danone? In the case of many companies that were widely acclaimed as innovative enterprises in their early years, it is difficult to know who the current senior executives are. A key responsibility of every senior management team has become the renewal of the cornerstones on which the corporate culture rests. I believe that the actions of senior executives (their values, principles and practices) when setting ambitious goals, supervising the progress towards those goals and facing the adversities that crop up on the road to innovation are fundamental in determining whether a culture regenerates itself or fades over time.

It is imperative that a company's senior executives understand and accept their fundamental responsibility in a process as critical as promoting innovation. The perception of employees makes it impossible to delegate any aspect of this responsibility.

ACKNOWLEDGEMENTS

My gratitude to Esteban Masifern, Carlos Cavallé and José Antonio Muñoz-Nájar for their suggestions on specific aspects which have substantially influenced the focus of this chapter;

to Dana Page and her IESE Gemba-2010 team for their contributions on Whole Foods Market;

to Edinardo Figueiredo and his Gemba-2010 team for their contributions on Pixar;

to Preeti Sharma and her Gemba-2010 team for their contributions on the Tata Group;

to Joost de Kinderen and his Gemba-2010 team and to Ariel Kestens and his Gemba-2009 team for their contributions on IKEA;

to Federico Ciardelli and his Gemba-2009 team for their contributions on Dyson;

to Andrey Lankovich and his Gemba-2008 team and to Manuel Lapeira and his Gemba-M-2010 team for their contributions on Procter & Gamble.

BIBLIOGRAPHY

CHRISTENSEN, C. M. (1999), "What Is an Organization's Culture?", Boston, MA: Harvard Business School Publishing, Note 9, pp. 399-104.

DESCHAMPS, J.-P. (2008), *Innovation Leaders: How Senior Executives Stimulate, Steer and Sustain Innovation,* New York, NY: John Wiley and Sons.

SCHEIN, E. (1988),*Organizational Culture and Leadership*, San Francisco, CA: Jossey-Bass Publishers.

SENGE, P. M. (1998),"The Practice of Innovation", *Leader to Leader* 9, Summer, pp.16-22, also in *Society for Organizational Learning*.

TUSHMAN, M. L., and C. A. O'REILLY III (2002), *Winning Through Innovation: A Practical Guide to Leading Organizational Change and Renewal*, Boston, MA: Harvard Business School Press.

VILÁ, J. (2008), "SettingUp Innovation Management: Roadmap towards a Fully Innovative Culture", *Paradigmes*, 0, Departament d'Universitats, Indústria i Empresa, Generalitat de Catalunya.

BBVA

Economic development revisited: How has innovation contributed towards easing poverty?

Manuel Mira Godinho
ISEG (Technical University of Lisbon), UECE

1. INTRODUCTION

The literature on the global distribution of income highlights how poverty worldwide has declined over recent decades. Sala-i-Martin (2006) summarized his findings by stating that poverty rates "in 2000 were between one-third and one-half of what they were in 1970 for all four [poverty] lines. There were between 250 and 500 million fewer poor in 2000 than in 1970". These findings have been confirmed by more recent research. Pinkovskiy and Sala-i-Martin (2009) pointed out that the percentage of the world population living on less than $1 a day (in PPP-adjusted 2000 dollars) declined from 26.8% in 1970 to 5.4% in 2006.

In sum, the message from the literature on the evolution of global poverty shows that, should the observed trends continue, *poverty* will probably be eradicated on planet earth by the middle of the present century. Such a message is of course good news. Particularly if we assume that underdevelopment is a synonym for extreme poverty in the developing world, we can readily accept that the observed decline in global poverty is highly correlated with progress towards economic prosperity.

However, as there are several methodological drawbacks to the way poverty estimates are carried out, one must be careful over their interpretation. The research quoted above refers to a poverty line set at $1 a day in contrast with the World Bank, which has been working with alternative poverty lines. International poverty lines are set close to the mean of the poverty lines identified in the poorest countries. For this reason, it is difficult to argue in favour of just a single poverty line. Over the last few years, the World Bank has mostly been working with a poverty line of $1.25 a day while also having recourse to four other alternative thresholds, $1, $1.45, $2 and $2.5 a day respectively (Chen and Ravallion, 2008). If one takes this last threshold, which is certainly adequate for the many developing countries that are not among the poorest, we quickly find that the 1990 figure of circa 3 billion of world poor did not undergo any decline through 2005. In any case, the available research does seem to confirm that poverty declined worldwide, where not in absolute numbers (even though much research confirms precisely that) at least in relative terms, as the world population has been continuously growing from about 5 billion in 1990 to almost 7 billion in 2010.

These observations of such a worldwide trend enable us to raise several important questions from an economic development standpoint. The central question relates to the possible causes behind such a positive trend. Certainly many factors are involved in accounting for what has been observed: rising literacy, better health coverage, an expansion in trade, the adoption of innovation and potentially- industrial policies and improvements to institutions being among the most important of such factors. In this chapter we are specifically interested in this wider context and discussion as to the role played by innovation in economic development. We will focus upon possible future roles for innovation in further pushing back global poverty and bringing about development worldwide in the next few decades.

Keeping these questions in mind, the present chapter unfolds into three sections. Section 2 provides an account of how the economic development literature has advanced and dealt with innovation, identifying what lessons one might derive from the many different perspectives that have emerged on the interrelationship between development and innovation. Section 3 then reviews the innovation literature and its approach to economic development from a rather symmetrical standpoint to the previous section. Finally, section 4 summarizes the main topics dealt with throughout the chapter and sets out a forward looking perspective on how innovation and development may be expected to interact over the next few decades.

2. HOW HAS THE ACADEMIC LITERATURE ON ECONOMIC DEVELOPMENT APPROACHED INNOVATION?

The early days

Despite being intimately inter-related in practice, innovation and development have to a large extent been dealt with as two separate academic topics. *Economic Development* emerged and evolved as an autonomous field of study essentially after World War 2, as the implementation of the Marshall plan in Europe and the newly-gained independence of the former European colonies raised the issue as to which were the best developmental policies for the new context. Regarding economic analysis of *Innovation*, many empirical studies on technological change and innovation have been carried out since the 1960s, in the wake of Solow's work, with the introduction of this topic as an autonomous subject in undergraduate or graduate curricula in effect since the 1980s.

For many years, the literature on economic development did not even mention the word innovation. As a matter of fact, the equivalent terms that dominated that literature until a few decades ago were technical progress or technological change. This is ironic as Joseph Schumpeter, best known for being the first academic to systematically conceive of innovation, published a book in 1912 entitled precisely *The Theory of Economic Development* (Schumpeter, 1912). In this work, he began by putting forward a model in which the economy operated in a circular flow. It was the introduction of innovations by pioneering entrepreneurs that enabled the economy to move out of the initial static equilibria of the circular flow. Indeed, it was also within this sequence that *creative destruction,* brought about by the introduction of radical innovations, generated the dynamics of

> “For many years, the literature on economic development did not even mention the word innovation. As a matter of fact, the equivalent terms that dominated that literature until a few decades ago were technical progress or technological change”

the business cycle. However this insight of Schumpeter was not carried over into studies on economic development produced in later decades, possibly as such a view was seen as more applicable to mature capitalist economies than to poorer developing nations.

Balanced versus unbalanced growth: searching for the development engine

Rather than being inspired by Schumpeter, early development theory was influenced by Keynesian growth models. In these earlier approaches, the capacity to raise saving levels so as to finance capital accumulation was seen as a key condition for growth to take place (Domar, 1946, Harrod, 1948). From this perspective, economic modernization and progress depended upon the possibility of raising savings and investment rates, an objective unachievable through policy measures. A similar emphasis on the accumulation of tangible capital was sustained by Marxist authors such as Dobb (1951).

Soon, however, there was a departure from these views, perceived as too simplistic in assuming the economy to be mono-sectorial. The debates rapidly moved on to the problems of balanced versus unbalanced growth and the structural composition of the economy. Lewis (1954) put forward his dualistic model of a two- sector economy, composed of a *traditional* and a *modern sector*. The *traditional sector* coincided with agriculture in rural areas while the *modern sector* essentially coincided with the modern industries concentrated in urban areas. Nowadays, we would state that the modern sector is the carrier of innovations into the economy. Trickle-down effects from the modern to the traditional sector were the expected drivers of modernization and economic development. The views on this sort of inter-sectoral dynamics were researched further by Hirshman (1958). According to this author, the critical aspect of development was not so much the saving and investment rates but the actual ability to mobilize entrepreneurial capabilities. In his view, existing entrepreneurs needed encouragement in order to concentrate their investments on specific sectors whose backward and forward linkages would generate leverage effects throughout the whole economy. This emphasis on the economic structure was later significantly expanded by perspectives focusing on the relevance of the international specialization of the different economies.

At odds with the *dualistic* and *structuralist* views, the proponents of *balanced growth* (Singer, 1952, Nurske, 1953) claimed that development required bringing about a coordinated expansion of several sectors. As markets are limited in developing economies and as overall output growth depends on existing demand, the balanced growth proponents stated that the existing sectors had to co-evolve to generate mutual demand large enough to provide the necessary leverage for overall economic growth. This was regarded as particularly important as developing countries were seen as having only limited opportunities for exporting to an international market dominated by the OECD economies.

It was this last insight that led to another set of inter-related approaches within the field of economic development. Prebisch (1950), based at the United Nations Economic Commission for Latin America, formulated his thesis that developing countries had to promote import-substitution policies as the world had evolved into a centre-periphery relationship in which developing nations were condemned to export raw materials and primary goods to rich nations, while importing capital goods and other technology-rich imports from the latter. Protectionism was needed to ensure that domestic markets could expand and exploit economies of scale—a critical characteristic of the most advanced technologies of the day. These views crystallized afterwards to form what became *dependency theory,* expounded upon by authors such as Furtado (1973), Frank (1975) and Amin (1973).

Which technologies are best for developing countries?

Since the 1970s there were, however, several signs indicating progressive discontent with the direction of development theory over the decades. Internally, there were those such as Seers (1969) who contested the dominating growth of fetishism, calling for academic analysis of development to turn instead to qualitative aspects more related to human needs. Amartya Sen became the leading exponent of this group of approaches, given his proposition that the critical aspect of development was the freedom of individuals to do or to become something. This right of access could only be achieved when individuals were equipped with capabilities appropriate for implementing their choices (Sen, 1980).

It was within this qualitative set of approaches that an influential book emerged entitled *Small is Beautiful* (Schumacher 1973). The book brought to the debates on development not only the idea that economic growth might not be a central objective, but further suggested that growth might be harmful and, perhaps best avoided by societies. These ideas originated within an intellectual climate rejecting the dominant prevailing materialistic culture of advanced capitalist economies, in conjunction with a growing global perception as to the limits of growth imposed both by environmental degradation and by limited stocks of non-renewable natural resources (Meadows *et al.,* 1972). In his book, Schumacher put forward the idea that many modern technologies were harmful and that societies had to gain by sticking to smaller-scale technologies, which might be either traditional, or possibly *intermediate, technologies*. Such *intermediate technologies* were portrayed as more productive than traditional technologies but with a lower capital intensity and much less damaging to the environment than modern scale-intensive technologies. Schumacher's ideas led to the establishment of the intermediate technology movement that branched out into two streams, one within a developing context, with many locally- based experiments with appropriate technology implemented over the years in poorer countries, and another within the developed economy context, with the search for environment-friendly technologies. It must be stated that Schumacher's work had very important earlier roots in the thinking of Mohandas Gandhi. India's independence leader had advocated small, local-based technology as a means for Indian labourers to become self-reliant and able to compete with the large-scale technologies deployed by the British. In fact, large scale technologies are typically centralized and, as such, were used by the colonial power as a way of concentrating production and imposing prices on indigenous populations.

This *intermediate* (or *appropriate) technology* perspective is convergent with Amartya Sen's perspectives on individual capabilities and self reliance. *Appropriate technology* is regarded as empowering the poor by allowing greater individual and local community autonomy while simultaneously respecting the environment. It was in keeping with these perspectives that radical views were advanced by thinkers such as Vandana Shiva (1992, 2000). In her 1992 book, she distilled criticisms made in respect of what is termed the *green revolution*. While many have defended the achievements brought about by the green revolution through the application of modern science to genetically recombining the existing varieties of agricultural species, many others have criticized it on social, political, health and environmental grounds. The green revolution represents a fascinating case-study for discussing how innovation impacts on development paths and how technological choice is a current problem that policies and societies should take into consideration.

It is interesting to note that the intermediate technology views have more recently diversified into quite a different perspective. C. K. Prahalad, who was well known for books on strategy and knowledge management, in 2004 published *The Fortune on the Bottom of the Pyramid: Eradicating Poverty through Profits*. The bottom of the pyramid (or simply BoP as it became known) is the 4 billion poor living worldwide on less than $2 a day. Basically, Prahalad's idea was to adapt and integrate the solutions of the past—development aid, subsidies, government support, exclusive reliance on deregulation and the privatization of public assets—within a broader market-based approach. He made a call "to mobilize the investment capacity of large firms with the knowledge and commitment of NGOs and the communities that need help" through the co-creation of unique solutions. Thus, in this view, the poor were not seen as a passive market upon which business firms impose existing products, but rather as an active part of the innovation process itself, which should involve multinational national corporations (MNCs) in the co-creation with them of new products adapted to their needs and wallets. Prahald's approach stimulated an important stream of literature focusing on: 1. how the poor should be involved in the co-creative process for their own benefit (e.g. Ramani *et al.,* 2009; Ghazi and Dusyters, 2009); 2. MNC corporate social responsibility in the Third World (e.g. Rangan *et al.*, 2007); and 3. specific case studies that show how MNCs profit from a global market worth $5 trillion, highlighting evidence on many important pro-poor innovations.[1] In a way, the literature on BoP innovation goes beyond the older literature on technology choice (e.g. Stewart, 1978), which tended to conceive of the option between endogenous (traditional) technology and foreign technology as alternative paths, as the newer approach emphasizes the integration of efforts made by the local poor in developing countries and (mostly foreign) MNCs.

[1] In this regard, the "Special Report on Innovation in Emerging markets" published by *The Economist* in its 17th April 2010 issue, is highly recommended as it provides many interesting examples of pro-poor (or by-the-poor) innovations.

Has development economics developed?

Apart from the recent (qualitative) contributions highlighted in the previous paragraphs, most research on development economics over the last two decades has displayed an analytical character, more concentrated on technical problems than on the actual challenges of development. This happened not only because the *old* development economics was discarded by mainstream economists as methodologically unsatisfactory, but also because the lack of advancement in developing countries helped to make policy makers there grow weary in the face of the excessively normative and impractical nature of the existing theories. As a result, in recent years, a significant percentage of research on development economics has followed a different route, especially under the auspices of the *new growth economics*, as a continuation and refinement of the economic growth models that Robert Solow and colleagues had put forward in the late 1950s and 1960s. Lundvall *et al.* (2009) provide an interesting account of this evolution, concluding that "currently mainstream economics tends to use developing countries' problems as offering interesting opportunities to make use of advanced theoretical models and econometric tools while the interest in understanding the structures that lie behind underdevelopment and the mechanisms that might trigger development tend to end up as being of secondary importance".

However, development economics has not exclusively moved along an analytical path. On the one hand, the perspectives opened up by Amartya Sen led to an important reconsideration of what exactly is meant by *development*, concentrating on the relevance of the freedoms and capabilities of both the individual and the society. On the other hand, much empirical work has been produced about successful instances of development and catch-up in recent decades, providing valuable insights into the strategies advanced for effectively learning and incorporating innovation into the development process. It is precisely to this last stream of literature that we turn in the next section of this chapter.

3. INNOVATION, LEARNING AND CATCHING UP: NEW PERSPECTIVES ON ECONOMIC DEVELOPMENT

This section offers a sort of mirror view of the previous one that reviews the research on *innovation* relating more directly to economic development. It starts by briefly presenting the main concepts of the *innovation literature* before highlighting the approaches to technology transfer and technological learning within a development context. Finally, it provides a summary of the approach to catching up, in which innovation is seen as a central aspect to countries attempting to swiftly move out of underdevelopment.

From innovation as a process to innovation as a system

Innovation has been defined as the first practical application of an invention. Normally, that application takes place in an organized market in which innovating firms introduce new products or supply already existing products through using new processes. As Fagerberg (2005) pointed out, "To be able to turn an invention into an innovation, a firm normally needs to combine several types of knowledge, capabilities, skills and resources. For instance, the firm may require production knowledge, skills and facilities, market knowledge, a well-functioning distribution system, sufficient financial resources and so on." In this view, innovation is essentially a knowledge-intensive process.

> "The forces behind innovation have been divided into two main groups of factors associated especially with *market opportunities* and *technological opportunities*"

The forces behind innovation have been divided into two main groups of factors associated especially with *market opportunities* and *technological opportunities*. The *market opportunities* argument was formalized into the so-called *demand-pull innovation model*. In this model, innovation is stimulated by needs such as illnesses or the search for more energy-efficient processes for which the market has yet to provide satisfactory solutions. This view was argued in a book, *Invention and Economic Growth* (Schmookler, 1966), based on the study of historical time-series of US patents, investment and production from around the turn of the 19th century through to the 20th century. Such a view came in for criticism from Mowery and Rosenberg (1979), who claimed that not all innovations stem from needs put forward in the marketplace. Specifically, these authors argued that many innovations, in particular in the industrial sectors that emerged in the second half of the 20th century such as electronics, tended to stem from the application of scientific discoveries or from the unexpected results of technological research and development. These innovations, driven by *technological opportunities,* led to the suggestion of a *science and technology-push model of innovation*. These two contrasting views on innovation came to be seen as complementary and later were integrated into the *interactive model of innovation* (Freeman, 1979) and elaborated in the recursive *chain-link innovation model* (Kline and Rosenberg, 1986).

It was in this theoretical sequence that systemic views of innovation emerged, suggesting the concept of an *innovation system*. The *innovation systems* literature (Freeman, 1987 and 1995; Lundvall, 1992; Nelson, 1993; Edquist, 2004; Malerba, 2002) has attempted to integrate the forces of demand with science and technology in a wider, systemic, context, in which different actors and institutions with a role in innovation interact. This view considers the introduction and adoption of innovations to be a complex process stemming from the coordination of efforts between a diversity of stakeholders. Furthermore, this approach has pointed out that the innovation process is strongly affected by historical trajectories and by normative environments -- i.e., it has demonstrated that innovation is an institutionally embedded process. The innovation systems approach builds upon not only the simpler models of the innovation process, which were briefly reviewed above, but also on the S&T system concept developed in the 1960s as well as the old and modern institutionalist schools.

In recent years, this innovation- systems approach has fed through to the analysis of economic development in two streams of analysis. Firstly, by putting forward the *national innovation system* concept,[2] which highlights the need for individual actors (firms, consumers, universities, financial operators, civil servants, intermediate

[2] An internet search carried out in mid-2010 of the "National Innovation System" concept, brought about 742,000 results in Google, while the search for the equivalent "National System of Innovation" expression brought about a further 266,000 results. Together, this represents more than 1 million references to the concept in documents available on-line!

organizations...) coordinating their efforts through collective strategies and forward-looking visions presented at the national level. This concept has been applied to a huge array of economies, initially to mature economies but increasingly to emerging economies and many poorer developing countries (for example, see Arocena and Sutz, 2000; Gu and Lundvall, 2006a and 2006b; Joseph 2006; Lastres and Cassiolato, 2005; Lastres, Cassiolato and Maciel, 2003; Liu and White, 2001; Oyelaran-Oyeyinka, 2006; Viotti, 2002). Secondly, as this approach has evolved into the analysis of *learning systems*, the focus has moved on to the mechanisms behind the production, adoption and diffusion of new productive knowledge, which are of course critical for developing economies. To a certain extent, this second stream of analysis connects both with the older literature on *technology transfer* as it looks into sources of technology within a development context and with the more recent literature on technological learning, both of which are briefly reviewed in the two subsequent points.

Technology Transfer

In the older literature on *technology transfer,* developing countries were portrayed as *followers, to a greater or lesser extent* expected to passively absorb and adopt innovations pioneered by leading economies. To this end, all they had to do was to tap into the sources of foreign technology, of which the most important was the import of capital goods.[3] Other channels consistently analyzed by the *technology transfer* literature included foreign direct investment (FDI), joint-ventures, technology licensing and subcontracting by original equipment manufacturers (OEMs). Concerning FDI, this has generally been seen as having a positive role in the transfer of know-how, though with significant limitations. A common view by the mid-1990s was that FDI was an effective means of transferring innovation but not necessarily the innovative capabilities themselves (Lall, 1996).

These technology transfer sources were extensively discussed in relation to the cases of several successful newly industrializing economies. Analyzing the four East Asian *dragons* (South Korea, Taiwan, Hong Kong and Singapore), Hobday (2000) pointed out how different mechanisms worked effectively in each case. In the case of South Korea, several of the Korean *chaebols* started by subcontracting production capacity as OEMs to large Japanese corporations before next proceeding to produce design and development while still supplying finished goods as subcontractors, before later entering into merchandising and selling their products under their own brands worldwide. This sequential process allowed them to absorb critical know-how from their contractors and acquire innovative capabilities in product and process engineering. In contrast to Korea, in the Taiwanese case the smaller electronics and IT firms of this country focused mostly on importing technology through licensing foreign technology, a path with a certain resemblance to that followed by Japan a few decades earlier (Freeman, 1987).

A relevant recent development in these perspectives on technology transfer has been the analysis of *global value chains* (GVCs) (Ernst, 2001; Kaplinsky, 2005). This analysis provides evidence and insight on how developing countries and their domestic companies are involved in global supply chains. A central topic of this perspective has been the governance mechanisms of the GVCs (Gereffi *et al.,* 2005), assessing with which statute and in which operations

[3] This view brings us back to the problem addressed by earlier development models: the ability of macroeconomic policy to optimize the savings and investment rates.

(assembly, design, marketing...) developing country companies participate in the GVCs.

It should be noted that the focus on the need and interest of absorbing foreign technology contrasts considerably with certain perspectives on economic development that for several decades influenced many countries, namely the *structuralist* and *dependency schools* that advised national governments and developing economy companies to rely as much as possible on their own resources and capabilities rather than on foreign know-how.[4]

Technological Learning

To a certain extent, the literature on technological learning in the developing world context provides an integration of the more interesting aspects of the apparently conflicting approaches that highlight the relevance of external and internal sources of technological development.

Technological learning has been defined as "any process by which the resources for generating and managing technical change (technological capabilities) are increased or strengthened" (Bell and Pavitt, 1993). This view portrays technology as much more than machinery-embodied knowledge and instead has focused on the cognitive aspects of the learning process, stating that "technology is a... bundle of knowledge, with much of it embodied in a wide range of different artefacts, people, procedures and organizational arrangements. These embodiments of knowledge include at least: product specifications and designs, materials and component specifications and properties; machinery and its range of operating characteristics; together with the various kinds of know-how, operating procedure and organizational arrangement needed to integrate these elements in an enormously variable range of different production systems" (Bell and Albu, 1999).

Much of this research on technological learning began by analyzing the mechanisms of technological accumulation at the individual firm level, specifically looking into large-scale companies from countries such as Argentina, Brazil, Mexico, South Korea and India (e.g. Dahlman and Fonseca, 1987; Katz, 1985). The focus was on how individual companies organized their process of capability- building through learning by doing, combined with endogenous R&D. More recently, however, the research on technological learning evolved into analysis of the more complex structures in which developing country companies interact with suppliers, customers and specialized knowledge-generating organizations, such as universities and R&D institutes. Thus, interest moved on from the realm of the individual company to analysis of *networks* or *clusters*. According to Bell and Albu (1999) such combinations of internally organized capabilities with external knowledge resources "have come to be described as *industrial innovations systems*, *technology systems* or *knowledge systems*".

In this systemic context, convergent to that put forward by the *innovation systems* literature, *technological learning* is seen as a dynamic process of capability acquisition and development, with the success of this process depending on both historical trajectories and on the institutional setting in which the process occurs. Furthermore, in contrast to the idea of technology absorption through the importing of capital goods, technological capabilities are not seen as the result of a single occurrence or event but as a time-dependent process associated with a long-term purposeful organization of efforts by

[4] One of the corollaries of these more inward-looking perspectives was that economic development had to be *balanced* with the simultaneous growth of all economic sectors as the developing countries could not overly rely on specialization and the opportunities presented by a trading system largely dominated by OECD countries.

firms and other relevant actors in the national innovation system.

Catching Up

Catching up refers to the ability of a given country to reduce its productivity differential vis-à-vis the leading economies over a given historical period (Fagerberg and Godinho, 2005). The catching-up literature has stressed that endogenous and exogenous factors combine in triggering productivity rises. The historical antecedents to the catching-up literature stretch back to early in the 20th century, with the work of Thorstein Veblen on the German catch up. However, more systematic contributions took place concomitantly with the early literature on development and economic growth, particularly the work of Gerschenkron (1962).

Gerschenkron adopted an essentially optimistic view about the possibilities of countries evolving out of underdevelopment, suggesting that the more backward a country is, the higher its potential for a fast catch-up process. This paradoxical view was seen as possible as underdevelopment originates a tension between the existing backwardness and the promises offered by economic development. Such tension would facilitate a quick rise in the investment rate and a concentration on the rising industries and technologies. This view of Gerschenkron has been extensively tested by many econometric studies, analyzing the negative relationship between initial GDP per capita and its rate of growth for samples with a large number of countries (e.g. Baumol, 1986; Lucas, 1988; Barro, 1991; Barro and Sala-i-Martin, 1992; Quah, 1993).

Despite its essentially optimistic outlook, for many in the developing world, the perspective on catching up has been negatively associated with linear historical views such as those put forward by W. W. Rostow (1960). Rostow stated that all countries have to evolve through pre-determined *stages of growth*, with developing nations expected to mature along similar lines to those of the US or the UK over the 19th and 20th centuries. According to Rostow, the economic development problem basically related to the capacity to mobilize the resources necessary for the *take-off* from backwardness to modernity. The rejection of Rostow's scheme to a certain extent echoes the arguments reviewed earlier about the need for developing countries to encourage and deploy appropriate technology. However, it should be noted that the research on catching up and the work of Gerschenkron himself do not impose any need for countries to evolve through similar stages of development with catching up linked to the conditions necessary for reaching (and eventually surpassing) the productivity levels of the best performing economies in each historical period in a relatively short period of time.

One possible reason for suspicious attitudes towards catching-up theory is the sheer frustration felt by both academics and policy makers in the developing world given the gigantic difficulties and backlashes faced in overcoming underdevelopment. However, it should be pointed out that there are certain variations within this approach ranging from the more positive views that accept the feasibility of *technological leapfrogging,* given certain *windows of opportunity* (e.g. Pérez and Soete, 1988), to other views that have emphasized the many existing barriers and a very diverse set of pre-requisites, especially in relation to the need for prior technological accumulation over lengthy periods of time (Pavitt, 1985).

The mainstream catching-up literature has focused precisely on these latter

> **“Some of the most influential empirical research on recent cases of catching up converges around findings within the framework put forward by economic historians working within an institutionalist perspective”**

aspects. Following Gerschenkron's insight, most research on catching up soon concentrated on the technological pre-requisites, viewing innovation as a central aspect to efforts by poorer economies to rapidly close the economic gap in relation to mature economies. It was as part of this research that a *technology gap* hypothesis was put forward (Posner, 1961; Fagerberg, 1987; Fagerberg and Verspagen, 2002), suggesting that the greater the technology gap, the greater the potential for catching up. However, while concentrating on the opportunities generated by investing in technology and innovation, the theory underlying this hypothesis has emphasized that technology is not a global, freely available public good. The difficulties faced in absorbing foreign technology gained particular emphasis in conjunction with the need to couple this with the local accumulation of technological know-how through endogenous R&D and other learning activities.

Furthermore, and also in accordance with Gerschenkron's insights and earlier approaches such as Veblen's (1915) analysis of the German catch-up, most research on *technology gaps* has stressed how catching-up candidates have to meet certain institutional pre-requisites. Abramovitz (1994) pointed out that aspiring economies should combine *technological congruence* with an indispensable *social capability*. By *technological congruence,* he meant the degree of coherence between economic aspects such as resource endowment, the degree of specialization in different technologies, the configuration of demand, the prevailing market characteristics and the country's position in terms of physical infrastructures. By *social capability,* he essentially encapsulated cultural and institutional factors, such as the levels of education and technical competence, the political climate, the capacity of business and research organizations to interact and, in more general terms, the economic culture framing entrepreneurship, innovativeness and the risk-propensity of economic actors.

This focus on *institutions* follows in the footsteps of work by economic historians such as Landes (1969, 1998) and North (1981, 1990). Landes argued that the earlier economic advancement of some European countries such as Britain was highly correlated with the presence of a set of efficient institutions, such as the enforcement of contracts and the personal liberties needed to guarantee geographic and social mobility. North adopted a standpoint closer to economics in the sense that he departs from the uncertainty that characterizes economic exchanges, depicting markets as part of the regulating institutions necessary for economic activity to advance. In his later work, he also portrays government as part of the institutional machinery needed to smooth out economic exchanges.

Some of the most influential empirical research on recent cases of catching up converges around findings within the framework put forward by economic historians working within an institutionalist perspective. Wade (1990) pointed out in his analysis of Taiwan and other East Asian countries that the appropriate combination of free market and government intervention accounts for the rapid industrialization achieved, particularly in terms of coordinated resource-allocation decisions. Rodrick (2007) was clear in stating that "the hallmark of development is structural change—the process of pulling the economy's resources from traditional low-productivity activities to modern high productivity activity", while also feeling the need to stress that "this is far from an automatic process and requires more than well-functioning markets. It is the responsibility of industrial policy to stimulate investments and entrepreneurship in new activities especially those in which the economy may end up having a comparative advantage".

Further to the focus on proper institutions and adequate coordination between markets and government, recent empirical literature has also demonstrated that rapidly catching up in terms of productivity typically results from a combination of selective protectionism along with opening up to foreign sources of knowledge (Chang, 2002; Hobday, 2000).

4. CONCLUDING REMARKS

One important topic present throughout the literature reviewed in the previous sections concerns the possibility of developing nations deciding on which technologies best fit their needs. As seen above, Rostow's linear view of all nations following a similar sequential trajectory imposed the fate of following in the steps of the leading economies on the developing world. In this perspective, the task of developing economies would above all be to concentrate on fine-tuning the absorptive mechanisms and sequentially adopt the technologies invented earlier on by the leading economies. Simplistically expressed, this is the rationale behind the most basic notions of technology transfer. It was in part the rejection of this view that led to the concept of *appropriate* (or *intermediate*) *technology* put forward by Schumacher and others, thereby suggesting there is a *technology space* from which many alternative picks are possible.

Currently, the literature on technological choice extends far beyond the developing countries context to claim that alternative technological routes might be pursued even in a developed economy context. Such a perspective, for example, underpins the longstanding critiques of Fordism and Taylorism that discuss the deskilling effects of modern technology (Braverman, 1974; Noble, 1977, 1984), the works of the Tavistock Institute on socio-technical systems, or the approach that proposes 'anthropocentric production systems' (Lehner, 1992).

The fact that some developing nations have been creating and effectively diffusing some appropriate (process and product) technologies seems to confirm the view that alternative technologies might actually be implemented with success. The recent introduction of the Tata Nano, designed and produced in India, is a very interesting example confirming this assertion. This new car has involved important process and product innovations that have been classified as radical and disruptive at a world level (Lim *et al.*, 2010).

However, despite the idea of a *technology space* -- within which different technological choices can be taken -- now being widely accepted, the consensus is also that such a *technology space* has no infinite possibilities given natural resource shortages and limited design alternatives. This view further implies that as a nation evolves towards higher GDP per capita levels, attempting to catch up with and eventually overcome the leading economies, the choices available within this *technology space* become much narrower. This happens because as a country or a regionally-concentrated cluster of firms begins to approach the state-of-the-art in any given technological field, the main constraint for advancing further becomes the scarcity of knowledge, which on the state-of-the-art frontier is complex and uncertain. The corollary is that when a nation seeks to become competitive at a world level in advanced technology, the practical choices for alternative technologies are significantly reduced. Of course, even in these circumstances, catching-up candidates do not need to invest in *narrow-space* high-technology across all economic activities, particularly in those activities whose outputs might be non-tradable. However, investing in at least a few of the most dynamic technologies of the day makes sense, not only because these typically generate higher earnings in expanding markets, but also because specialization in such technologies might produce the network, trickle-down and dynamic effects mentioned by the earlier development literature (e.g. Rosenstein-Rodan, 1943; Lewis, 1954; Hirshman, 1958, or Kaldor, 1966).

Another important topic dealt with by the literature reviewed in the sections above relates to the question of knowing whether a 'proper' or 'most advisable' sectoral composition of the economy is desirable. We may easily understand that most of the arguments set out in the paragraph above in relation to the choice of technologies and high tech investment can easily be replicated in the context of this discussion on the sectoral composition of the economy. Balanced development might certainly be more desirable in social terms as it would avoid mass migrations or the high unemployment costs stemming from swift changes in the economy's composition. Nevertheless, as economies wish to move on to leading edge technologies, there is hardly any alternative to accepting the roller coaster of creative destruction, at least to the extent that the world continues to be a collection of competitive nations and regions as has happened over the last millennium. Innovation is the epicentre of Schumpeterian dynamics and even though all sectors perform innovative activities, the intensity of such activities is unevenly distributed across sectors.

The catching-up literature has shown that the nations that have been the most successful in advancing rapidly in economic terms are those which have specialized in certain technologies and sectors. As pointed out by Fagerberg and Godinho (2005), the existing empirical evidence confirms that "the countries that have been most successful in catching up, namely South Korea, Taiwan, and Singapore (and Japan before them) have all—after initially having acquired some capabilities through more traditional activities—aggressively targeted the most technologically progressive industries of the day, in which they today play an important role". The higher complexity of the newer technologies in each historical period opens up the possibility for innovative firms and countries to capture niche markets with potential monopolistic rents. Further to these arguments that focus on supply-side aspects, on demand grounds the analysis has also shown that nations have advantages in specializing in certain sectors rather than in others. Following on from earlier insights into the technology-gap theory (Posner, 1961) and the life cycle approach (Vernon, 1966), Lafay (1982) precisely demonstrated that nations specializing in the products and sectors where international demand grows faster also turn in better economic growth performances.

The advocacy above of the advantages of certain sectoral specializations over others does not mean, however, that an 'optimal' specialization might exist for each given historical period. Specialization should be seen and treated as context-dependent. Resource allocation is a primary determinant of specialization, as pointed out by classical trade theory. For example, economies well endowed with resources such as beaches, sun, forests or valuable monumental heritage have advantages in specializing in tourism-related services. Indeed, the geography and characteristics of each country, in terms of its territory, population and market size, bear important implications in terms of possible sectoral specializations. However, historical evidence has made it clear that (at least for larger-sized) developing nations willing to succeed, there is little alternative but to invest in the most dynamic and innovative technologies and industries.

Another central lesson stemming from the literatures reviewed is that the successful adoption and creation of innovations in the developing world needs a corresponding institutional climate; otherwise investment in technological accumulation risks failing just as investment in capital goods or infrastructures proved a failure in previous development contexts. Taking as a basis their historical reality, countries need to build upon and adapt their institutions to the challenges of dealing with technological knowledge that often has a scientific background. This means a capacity for organizing and strengthening national innovation systems, setting up and improving communication channels between the relevant actors, while simultaneously nurturing the necessary trust so that these interactions also intensify in quantity and quality.

The recent advances of countries like China, India or Brazil stem to a large extent from their ability to put in place the building blocks of their national innovation systems. For smaller countries, however, this task may prove harder as they lack the capacity for similarly establishing critical masses of resources and benefiting from dynamic economies of scale and networking effects to effectively gain a proper return on their investments. Smaller countries might, however, adopt different technological strategies through relying more on natural resources or service-oriented strategies rather than on classical industrialization strategies, while simultaneously integrating further into the global knowledge networks so as to screen the technology sources needed.

One aspect that has changed dramatically over the last two decades in relation to innovation and economic development has been the global geography of business R&D. While twenty years ago the R&D carried out by MNCs was concentrated almost exclusively in their home countries, the situation has now substantially changed with a degree of R&D delocalization to third countries never before experienced. Another aspect that has also changed dramatically has been the international organization of

> “This means a capacity for organizing and strengthening national innovation systems, setting up and improving communication channels between the relevant actors, while simultaneously nurturing the necessary trust so that these interactions also intensify in quantity and quality”

intellectual property right (IPR) systems. The TRIPS agreement was introduced as an Annex to the founding treaty of the WTO, and as such almost all the world has come (or is coming) under a common set of rules for IPR. It is interesting to note in this respect how countries like India and Brazil (and to a lesser extent China as well), which disputed several provisions of that agreement, are now among the countries where domestic IPR usage is rising fastest. As noted by Godinho and Ferreira (2010), “both China and India have been experiencing a historical take-off in the use of intellectual property rights (IPR). As for national IP office applications, the evidence is that by 2009 China became number one worldwide in trademark applications, while India is just behind the US, Japan and the Republic of Korea. Concerning patent filings, China ranks third worldwide and India ranks ninth.” Brazil is also more intensely using trademarks and patents and while for this latter IPR type it is not yet in the global top 10, it ranked third worldwide for the former at the end of 2009.

The trends in this sort of indicators point to two different aspects that should be kept in mind for future debate on innovation and development. The first is that innovation is clearly becoming a central part of emerging-economy development processes in the same way as happened before in other cases of successful catch up. The second is that, nowadays, developing economies can hardly look inward if they want to further their economic development prospects. On the contrary, while caring for domestic conditions, they need to search thoroughly for adequate sources of know-how, learn to benefit from participation in knowledge networks, compete for outward FDI in R&D and adapt creatively to the complexities of global institutional frameworks, such as IPR.

Of course, as innovation becomes a central component of economic development, as is happening in China, India and other emerging economies, developing countries will need wise policies to deal with both environmental spillovers and the Schumpeterian waves of *creative destruction*. Research carried out over the past decade shows precisely how the acceleration in innovation has increased income concentration in the developed economies since the 1980s (Levy and Murnane, 2007). Similar effects might be expected in the developing economies as well if appropriate policies are not implemented to combine leading-edge innovation with what has been termed *pro-poor innovation*.

REFERENCES

Abramovitz, M. (1986), "Catching up, forging ahead, and falling behind", *Journal of Economic History* 46(2), pp. 385-406.

Amin, S. (1973), *Le Développement inégal: essai sur les formations sociales du capitalisme periphérique,* Paris: Editions de Minuit.

Arocena, R., and J. Sutz (2000), "Looking at National Systems of Innovation from the South", *Industry and Innovation*, 7(1), pp. 55-75.

Barro, R. J. (1991), "Economic Growth in a Cross Section of Countries", *The Quarterly Journal of Economics* 106(2), pp. 407-443.

Barro, R. J., and X. Sala-i-Martin (1992). "Convergence", *Journal of Political Economy* 100(2), pp. 223-251.

Baumol, W. J. (1986), "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show, *American Economic Review* 76(5), pp. 1072-1085.

Bell, M., and M. Albu (1999), "Knowledge Systems and Technological Dynamism in Industrial Clusters in Developing Countries", *World Development* 27(9), pp. 1715-1734.

Bell, M., and K. Pavitt (1993), "Technological Accumulation and Industrial Growth: Contrasts between Developed and Developing Countries", *Industrial and Corporate Change* 2(2), pp.7-209.

Braverman, H. (1974), *Labor and Monopoly Capital*, New York: Monthly Review Press.

Chang, H.-J. (2002), *Kicking Away the Ladder: Development Strategy in Historical Perspective*, London: Anthem Press.

Chen, S., and M. Ravallion (2008),*The World is Poorer than We Thought, but No Less Successful in the Fight Against Poverty,* World Bank, Policy Research Working Paper 4703.

Dahlman, C. J., and F. V. Fonseca (1987), *From Technological Dependence to Technological Development: The case of Usiminas Steel Plant in Brazil,* Economic Commission for Latin America, Buenos Aires.

Dobb, M. (1951), "Some Aspects of Economic Development: Three Lectures", Delhi School of Economics, Ranjit Publishers.

Domar, E. (1946), "Capital Expansion, Rate of Growth, and Employment", *Econometrica* 14, pp. 137-47.

Edquist, C. (2004), "Systems of Innovation—Perspectives and Challenges", in J. Fagerberg, D. Mowery and R. Nelson (eds.), *The Oxford Handbook of Innovation,* Oxford: Oxford University Press.

Ernst, D. (2001), *Global Production Networks and Industrial Upgrading—A Knowledge-Centered Approach* Economics Study Area Working Papers 25, East-West Center, Economics Study Area.

Fagerberg, J. (1987), "A Technology Gap Approach to Why Growth Rates Differ", *Research Policy* 16, pp. 87-99.

Fagerberg, J. (2005), "Innovation: A Guide to the Literature", in J. Fagerberg, D. Mowery and R. Nelson (eds.), *The Oxford Handbook of Innovation,* Oxford: Oxford University Press.

Fagerberg, J., and M. M. Godinho (2005), "Innovation and Catching-Up", in J. Fagerberg, D. Mowery and R. Nelson (eds.), *The Oxford Handbook of Innovation,* Oxford: Oxford University Press.

Fagerberg, J., and B. Verspagen (2002), "Technology-Gaps, Innovation-Diffusion and Transformation: an Evolutionary Interpretation," *Research Policy* 31, 8-9, pp.1291-1304.

Frank, A. G. (1975), *On Capitalist Underdevelopment,* Oxford: Oxford University Press.

Freeman, C. (1979) "The Determinants of Innovation: Market Demand, Technology and the Response to Social Problems", *Futures* 11(3), pp. 206-215.

Freeman, C. (1987), *Technology, Policy, and Economic Performance: Lessons from Japan,* London, New York: Pinter Publishers.

Freeman, C. (1995), "The 'National System of Innovation' in Historical Perspective", *Cambridge Journal of Economics* 19(1), pp. 5-24.

Furtado, C. (1973), "The Concept of External Dependence in the Study of Underdevelopment", in Wilber (ed.), *Political Economy of Development and Underdevelopment*.

Gereffi, G., J. Humphrey, and T. Sturgeon (2005), "The Governance of Global Value Chains", *Review of International Political Economy* 12(1).

Gerschenkron, A. (1962), *Economic Backwardness in Historical Perspective,* Cambridge, MA: The Belknap Press

Ghazi, S. S., and G. Dusyters (2009), "Serving Low-Income Markets: Rethinking MNCs Strategies", in G. Duysters, W. Dolfsma and I. Costa (eds.), *Multinationals and Emerging Economies: the Quest for Innovation and Sustainability,* London: Edward Elgar.

Godinho, M. M., and V. Ferreira (2010), "Analyzing the Evidence of an IPR Take-off in China and India", paper presented to the *7th Asialics International Conference on Global Recession and Reform of Innovation Systems in Asia*, 15-17 April, Taipei, Taiwan.

Gu, S., and B.-Å. Lundvall (2006a), "Policy Learning as a Key Process in the Transformation of the Chinese Innovation Systems", in B.-Å. Lundvall, P. Intarakurmend and J. Vang (eds.), *Asian Innovation Systems in Transition,* Edward Elgar.

Gu, S., and B.-Å. Lundvall (2006b), "China's Innovation System and the Move Toward Harmonious Growth and Endogenous Innovation", *Innovation, Management, Policy and Practice* 8 (1/2), pp. 1-26.

Harrod, R. F. (1948), *Toward a Dynamic Economics: Some Recent Developments of Economic Theory and Their Application to Policy,* London: Macmillan.

Hirshman, A. O. (1958), *The Strategy of Economic Development,* New Haven, CO: Yale University Press.

Hobday, M. (2000), "East versus Southeast Asian Innovation Systems: Comparing OEM- and TNC-led Growth in Electronics", in L. Kim and R. Nelson (eds.), *Technology, Learning & Innovation: Experiences of Newly- Industrializing Economies,* Cambridge: Cambridge University Press.

Joseph, K. J. (2006), *Information Technology, Innovation System and Trade Regime in Developing Countries: India and the ASEAN*, London and New York: Palgrave Macmillan.

Kaldor, N. (1966), *Causes of the Slow Rate of Economic Growth of the United Kingdom: An Inaugural Lecture,* London: Cambridge University Press.

Kaplinsky, R. (2005), *Globalization, Poverty and Inequality: Between a Rock and A. Hard Place,* Cambridge: Polity.

Katz, J. (1985), "Domestic Technology Innovations and Dynamic Comparative Advantages: Further reflections on a Comparative Case-Study Program", in N. Rosenberg and C. Frishtack (eds.), *International Technology Transfer: Concepts, Measures and Comparisons,* New York: Praeger.

Kline, J., and N. Rosenberg (1986), "An Overview of Innovation", in R. Landau and N. Rosenberg (eds.), *The Positive Sum Strategies: Harnessing Technology for Economic Growth,* Washington DC: National Academic Press.

Lafay, G. (1982), "Stratégies de specialisation ou division internationale du travail", in J. Reiffers (ed.), *Économie et finance internationals,* Paris: Dunod.

Lall, S. (1996), *Learning from the Asian Tigers,* Basingstoke: Macmillan.

Landes, D. (1969), *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present,* Cambridge University Press.

Landes, D. (1998), "Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor", W. W. Norton.

Lastres, H., and J. Cassiolato (2005), "Innovation Systems and Local Productive Arrangements: New Strategies to Promote the Generation, Acquisition and Diffusion of Knowledge in Innovation and Economic Development", Special issue of *Innovation: Management, Policy & Practice* 7(2).

Lastres, H., J. Cassiolato, M. Maciel (2003), "Systems of Innovation for Development in the Knowledge Era", in J. Cassiolato, H. Lastres and M. Maciel (eds.), *Systems of Innovation and Development: Evidence from Brazil,* Cheltenham: Elgar, 2003.

Lehner, F. (1992), *Anthropocentric Production Systems: The European Response to Advanced Manufacturing and Globalization,* Brussels: CEC, DGXII-Monitor/FAST.

Levy, F., and R. J. Murnane (2007), "How Computerised Work and Globalisation Shape Human Skill Demands", in M. M. Suárez-Orozco (ed.), *Learning in the Global Era: International Perspectives on Globalization and Education*, Berkeley, CA: University of California Press.

Lewis, A. (1954), "Economic Development with Unlimited Supplies of Labour", *Manchester School of Economic and Social Studies* 22, pp. 139-191.

Lim, C., S.-H. Han and H. Ito (2009), "Disruptive Innovation by a Developing- Country Firm: Product Development Processes without Strong Technological Prowess in Tata Motor's Nano",

paper presented to the *7th Asialics International Conference on Global Recession and Reform of Innovation Systems in Asia*, 15-17 April, Taipei, Taiwan.
Liu, X., and S. White (2001), "Comparing Innovation Systems: a Framework and Application to China's Transitional Context", *Research Policy* 30(7), pp. 1091-1114.
Lucas, R. Jr. (1988), "On the Mechanics of Economic Development", *Journal of Monetary Economics* 22(1), pp. 3-42.
Lundvall, B.-Å. (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning,* London: Pinter.
Lundvall, B.-Å., *et al.* (2009), "Bridging Innovation System Research and Development Studies: Challenges and Research Opportunities", paper presented to the *6th Globelics International Conference,* Dakar, Senegal: 6-8 October.
Malerba, F. (2002), "Sectoral Systems of Innovation and Production", *Research Policy* 31(2), pp. 247-264.
Meadows, D. H. *et al*. (1972), *Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind,* Universe Books.
Mowery, D. and N. Rosenberg (1979), "The Influence of Market Demand upon Innovation: a Critical Review of some Empirical Studies", *Research Policy* 8, pp. 102-153.
Nelson, R. R. (1993), *National Systems of Innovation: A Comparative Analysis,* Oxford: Oxford University Press.
Noble, D. (1977), *America by Design,* New York: Knopf.
North, D. (1981), *Structure and Change in Economic History,* Norton.
North, D. (1990), *Institutions, Institutional Change and Economic Performance,* Cambridge University Press.
Nurkse, R. (1953), *Problems of Capital Formation in Underdeveloped Countries,* Oxford: Blackwell.
Oyelaran-Oyeyinka, B. (2006), "Systems of Innovation and Underdevelopment: An Institutional Perspective", *Science, Technology and Society* 11(2), p. 239.
Pavitt, K. (1985), "Technology Transfer among the Industrially-Advanced Countries: An Overview", in N. Rosenberg and C. Fristack (eds.), *International Technology Transfer: Concepts, Measures and Comparisons*. New York: Praeger.
Perez, C., and L. Soete (1988), "Catching up in Technology", in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds.), *Technical Change and Economic Theory,* London. Pinter.
Pinkovskiy, M., and X. Sala-i-Martin (2009), *Parametric Estimations of the World Distribution of Income*, NBER Working Paper 15433.
Posner, M. (1961), "International Trade and Technological Change", *Oxford Economic Papers* 13, pp. 323-341.
Prahalad, C. K. (2004), *The Fortune on the Bottom of the Pyramid: Eradicating Poverty Through Profits*, Upper Saddle River, NJ: Wharton School.
Prebisch, R. (1950). *The Economic Development of Latin America and Some of its Problems,* New York: ECLA/CEPAL.
Ramani, S., S. Sadre Ghazi and G. Duysters (2009), "Taking Innovation to the Poor: Lessons from Sanitation Activists in India", paper presented to the *VIth Globelics International Conference,* Dakar, Senegal: 6-8 October.
Quah, D. (1993), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review* 37(2-3), pp. 426-434.
Rangan, V. K. *et al.* (2007), *Business Solutions for the Global Poor: Creating Social and Economic Value,* John Wiley & Sons, Inc.
Rodrick, D. (2007), *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth,* Princeton: Princeton University Press.
Rosenstein-Rodan, P. N. (1943), "The Problems of Industrialisation of Eastern and South-Eastern Europe", *The Economic Journal*.
Rostow, W. W. (1960), *The Stages of Economic Growth,* Cambridge: Cambridge University Press.
Sala-i-Martin, X. (2006), "The World Distribution of Income: Falling Poverty and... Convergence, Period", *Quarterly Journal of Economics* 2, pp. 351-397.
Schmookler, J. (1966), *Invention and Economic Growth,* Harvard University Press.
Schumacher, E. F. (1973), *Small Is Beautiful: A Study of Economics as if People* Mattered, Vintage.
Schumpeter, J. A. (1912), *The Theory of Economic Development.* Leipzig, Duncker and Humblot, translated by R. Opie, Harvard University Press, 1934, Reprint, Oxford University Press, 1961.
Seers, D. (1969), "The Meaning of Development International", *Development Review* 11(4), pp. 3-4.
Sen, A. (1980), "Equality of What?", *The Tanner Lecture on Human Values*, vol. I, Cambridge: Cambridge University Press, pp. 197-220.
Shiva, V. (1992), *The Violence of the Green Revolution: Third World Agriculture, Ecology and Politics,* New Jersey: Zed. Books Ltd.
Shiva, V. (2000), *Stolen Harvest: The Hijacking of the Global Food Supply,* South End Press.
Singer, H. W. (1958 [1952]), "The Mechanics of Economic Development", in A. N. Agarwala and S. P. Singh (eds.), *The Economics of Underdevelopment,* London, New York: Oxford University Press, 1973.
Stewart, F. (1978), *Technology and Underdevelopment,* London: Macmillan,
Veblen, T. (1915), *Imperial Germany and the Industrial Revolution,* New York: Macmillan.
Vernon, R. (1966), "International Investment and International Trade in the Product Cycle", *Quarterly Journal of Economics* 80(2).
Viotti, E. (2002), 'National Learning Systems: A New Approach on Technological Change in Late Industrialising Economies and Evidence from the cases of Brazil and South Korea", *Technological Forecasting & Social Change* 69, pp. 653-680
Wade, R. (1990), *Governing the Market: Economic Theory and the Role of Government in East-Asian Industrialization,* Princeton: Princeton University Press.

Innovation and the Service Economy

Ian Miles
Manchester Business School,
University of Manchester, U.K.

This essay will be concerned both with Innovation in Services and Service Innovation. *Innovation in Services refers* to both product and process innovation in service firms, sectors and industries. It may involve the development of new or improved services, while *Service Innovation* specifically relates to the creation of new services—though this can involve service-providing organisations in all sectors. The phrases have these different meanings first, because the term "service" can refer to the service product or to the industries that specialise in such products; and second, because service products can be generated by organisations that do not specialise in services. Thus manufacturing firms may deliver all sorts of customer and after-sales services.

Both terms reflect themes that were long neglected by social and management researchers, but that have received increasing attention since the 1980s and especially since the 1990s. Table 1 displays data on the prevalence of various terms in publication titles. These results are certainly "noisy", but point to a striking increase in attention to these themes. (The dramatic increase in use of "service innovation" in the most recent period partly reflects use of this phrase in library service research and a range of computer and informatics contexts.) While innovation studies were taking off—almost 300 publications using "innovation" in their titles were recorded in 1970 alone, and almost 500 in 1975 alone—the overwhelming focus was on innovation in manufacturing industries. We can locate some earlier studies of innovation in health and local government, for example, but there was rarely much attention to services as a whole, or even to a broad range of types of service industry. Manufacturing industry provided the paradigm for innovation analysis, and services would thus be considered as deviant, low-innovation industries. Where there were technology-based services—such as telecommunications or indeed health—then the main source of innovation was seen to come from manufacturing industries such as electronics or pharmaceuticals. Service industries and organisations were largely passive receivers of these innovations. Thus in Pavitt's (1984) classification of industrial innovation styles, service industries were all classified as "supplier-led"—though he had modified this view by the 1990s (Pavitt, 1994).

The growth of attention to innovation involving services has been driven by the rising significance of service activities in industrial societies and around the world, as well as the emphasis on service in the competitiveness of firms of all types. Recent research draws on many sources, including earlier traditions of work on service organisation and marketing, as will be indicated below. Some researchers stress the continuities between innovation in service and manufacturing industries (this is particularly the case with research using large-scale surveys) while others stress distinctive features of innovation in services (this is particularly prevalent in case study research). Two rather similar accounts of these various approaches were developed by Gallouj (1998) and Coombs and Miles (2000). Each account proposes that approaches to service innovation can be conveniently classified into three groups (they even give the same name to two of these groups). Each of these classifications has been adopted by several subsequent scholars, but in a recent overview of service innovation research Droege et al (2009) proposed that the two classifications actually propose four approaches in all:

- Assimilation approaches (noted by Coombs and Miles)—The basic idea here is that most economic attributes of services are fundamentally similar to those of manufacturing sectors. What differences exist are more a matter of (often relatively minor) quantitative placement on one or other continuum. Both services and manufacturing can be effectively studied and statistically documented according to the methods and concepts developed for manufacturing. Such approaches assume that the theories and concepts developed in manufacturing contexts readily apply to innovation in services. Innovation can be measured in similar ways, and is liable to be produced and managed in similar ways. What differences there are reflect the fact that services tend to lag behind other sectors. Such an approach is apparent in many of the earlier statistical studies of innovation in services that deployed the data produced in the Community Innovation

Table 1. Publications located by using various search terms

YEARS	Terms sought for in document titles			
	Innovation in service	Innovation in services	Service innovation	New service development
1975-1979	0	0	1	0
1980-1984	2	0	5	1
1985-1989	3	6	2	9
1990-1994	5	5	4	6
1995-1999	12	45	20	12
2000-2004	24	92	83	69
2005-2009	57	99	417	81

Source: data produced by using terms for various time periods in Harzing's *Publish or Perish* (Harzing, 2010), searching all types of publications and examining title words only. An effort has been made to remove duplications by examining document authors and titles—while this had a major impact on a few cases (years with few publications, in particular), the overall trends are unaffected. There is some overlap between cases in the various columns, sometimes reflecting more than one of the terms being used; sometimes the search tool simply fails to differentiate between the terms. Because the term "service innovation" frequently received hits where formulations beginning "innovation in..." were used instead, the data in the fourth column come with the word "in" barred from the title.

Surveys (CIS). Such studies typically failed to reveal striking differences in the ways manufacturing and service firms set about innovating. A similar viewpoint is advocated in many mainstream accounts of topics such as trade and productivity, where it is suggested that existing instruments will work effectively to describe the service economy.

- Technologist approaches (identified by Gallouj. In Gallouj and Savona (2010) he suggests that these are actually the same as Coombs and Miles' assimilation approach, though Droege *et al*. consider the two approaches to be distinctive). Here stress is put on the important role of new technologies (especially Information technologies) in services. Gallouj and Savona see this as leading to an assimilation of ideas from studies of innovation in manufacturing, which also tend to stress technological innovation. But some authors have stressed technological innovation, while arguing that the trajectory of service innovation is distinctive. For example, the "reverse product cycle" proposed by Barras (1986, 1990) implies that service organisations follow a distinctive trajectory of technology-based innovation, beginning with use of new technology to render production of services more efficient, and culminating in the creation of new services. The emphasis on technology may resemble that of many assimilationists, but the upshot is more like the demarcation approach we discuss next.
- Demarcation approaches (identified by both sets of authors)—argue that service activities are highly distinctive. They may still be poorly understood, but what is clear is that in many respects they have dynamics and features that require novel theories and instruments. This approach is displayed in many case studies of service activities. It suggests at one extreme that quite new instruments are required for investigation of service activities, or that the results of established instruments need to be interpreted in new ways. For instance, since services conduct little R&D (on the whole), R&D-intensity is a poor indicator for identifying "high-tech" or "knowledge-intensive" services, and new approaches are required (e.g. skill profiles of the workforce). Since much service internationalisation takes the form of investment, franchising and partnerships rather than conventional exports, the analysis of services "trade" has to pay more attention to such modes of presence. The distinctive features of services include intangible and unstoreable products, and high degrees of interaction with customers (up to the point where consumers are often seen as "coproducing" services). Such features not only mean that service industries lag behind In terms of innovation, but also that their types of innovation and innovation-management processes are very different from those seen in manufacturing. A case for demarcation is also made in much of the service marketing literature, and in some studies of productivity analysis that point to particular problems in assessing service productivity in conventional terms -- e.g. Gadrey, 2002, (Grönroosa & Ojasalo, 2004)
- Synthesis approaches (proposed by both sets of authors)—accept that studies of services bring to the fore issues that require examination. But the idea here is that these are not exclusive to service industries and organisations. Thus studies of service innovation have

highlighted features of innovation that have been neglected in most examination of manufacturing innovation, and the argument goes that a comprehensive analysis and more adequate indicators can provide an enriched understanding of innovation right across the economy. This will not only cover the service activities of manufacturing firms, but also help account for variations within and across goods and service innovation.

The idea that a synthesis of approaches to innovation in manufacturing and services can be achieved is a promising one. For one thing, many manufacturing firms actually sell services as well as goods, and all of them produce some services for internal or external use. It is likely that innovation in these service activities will differ from conventional manufacturing product and process innovations; for example, it is likely that the web portal of a manufacturing firm will develop along similar lines to, and pose similar issues to, that of a service firm.

Furthermore, it can be argued that in many respects there is convergence between manufacturing and service sectors (Miles, 1993). One aspect of this convergence is that there is a greater resemblance between manufacturing and the traditional view of services (for example, producing more customised products, having closer links with consumers, etc.). At the same time, many services are becoming like traditional manufacturing (standardised and mass production of services by large firms, for example).

Another aspect of convergence may be the increased emphasis on service on the part of manufacturers. Thus Howells (2001) is just one of many recent researchers who have studied the "servicisation" of manufacturing (and extractive) firms. (For a survey-based study, see Avadikyan and Lhuillery, 2007, and an examination of small and medium-sized firms' goods and service strategies by Susman *et al.*, 2006). More generally, there have also been many accounts of servicisation/ servitisation processes in recent years.) Typically, this involves providing services related to the goods manufacturers produce, or to their production processes. In the former case, the new services may be "product services" such as after-sales support, or other ways of redefining the product that is sold to include, or even to consist of, services, rather than merely the delivery of a material artefact. Sometimes servicisation involves complementing the goods with services such as finance, insurance, maintenance, software, etc. Sometimes it involves a shift to a service focus, in which the firm sees its job in terms of providing the outcomes for customers that the goods themselves would be used to create: the firm can then sell a promised amount of service rather than sell—or even rent—the goods. A famous case of this is Rolls-Royce contracting to supply hours of flight time rather than aircraft engines; and the efforts by computer companies to sell cloud-computing services rather than a computer kit itself can be seen in a similar light. Such servicisation strategies are liable to influence innovation pathways, as different costs are internalised and externalised by the partners. The manufacturer will need to pay more attention to the ways in which its goods are consumed—for example, by monitoring usage through new sensors and software—and in turn this might promote new product services in providing customer support and equipment maintenance and disposal.

Even without the phenomena of convergence and servicisation, the synthesis

> "The manufacturer will need to pay more attention to the ways in which its goods are consumed—for example, by monitoring usage through new sensors and software—and in turn this might promote new product services in providing customer support and equipment maintenance and disposal"

approach would argue for comparative studies of (various) manufacturing and service sectors, and examination of the service activities of manufacturing sectors; it certainly does not imply that there is no need for close examination of innovation in services and service innovation. Rather, the issues raised in such studies should be viewed in terms of their potential importance across the whole economy.

SERVICES: DIVERSITY AND COMMONALITIES

Just as service innovation was long neglected in innovation studies, so the service sectors were long neglected in economic analysis and, not least, in the development of economic statistics. For a long time indeed, there was very limited information available on the "tertiary sector" (which was even sometimes known as the "residual sector"). Even now statistical data are often sparse, though this situation is being addressed very seriously by statisticians in many countries and international organisations. One achievement has been to establish a far more detailed classification of service industries than was available previously, and Table 2 outlines the current high-level structure of the standard industrial classifications (NACE Rev 2), in which service industries feature as sections G to R.[1] ("Section" is used by statisticians as less ambiguous than "sector".)

This statistical classification demonstrates the range of activities that are covered by service sections. Some services store, transport and repair goods—and indeed catering services can also produce meals from raw ingredients. Some services deal directly with people—educating or healing them, providing haircuts and other personal services. Some are much more concerned with processing information—moving it around as in telecommunications services, creating new knowledge as in research services, and applying knowledge for business or personal use as in professional services.

This wide range of activities already suggests that we might find different sorts of innovation taking place in the various sectors: surgical or pharmaceutical innovations may be important for hospitals but not for supermarkets or hotels; new financial products may have little relevance for sports centres or garages, and so on. These different sections are engaged in very different sorts of activity, and may thus be undertaking quite different sorts of innovation—some supplier-led, perhaps, while others may be much more the products of the firms themselves. Additionally, there are important differences in terms of the way in which

[1] Section T actually includes the role of households as employers of domestic personnel, which was traditionally a major type of service employment—"domestic service", household services.

the sections are typically organised. Many sections are dominated by smaller firms than is typical for manufacturing, and indeed there are many micro-businesses, involving just a few employees, in many services—family shops, freelance artists, consultants and accountants. But some sections are dominated by larger organisations—the financial services are typically composed of larger firms, and public services like health and education can be immense—the UK's National Health Service employed practically 1.5 million people in 2010! The occupational profile of sections also varies widely—some sectors have high levels of unskilled employees, while others are the most knowledge-intensive ones in the economy (in terms of educational credentials, at least). In innovation research, particular attention has been paid to two of the latter areas of service activity—public services (NACE sections O, P and Q), and Knowledge-Intensive Business Services (KIBS, mainly NACE section M).

It is interesting to note, in contrast, that earlier explanations of the supposedly poor productivity growth in services related this to the low-skill nature of many of the industries (e.g. Fuchs, 1968). There are also differences in terms of the markets served—consumers, businesses, and public authorities. (For documentation of these variations, see Miles, 2008).

Table 2. Broad Structure of NACE Rev. 2 (NACE stands for "Nomenclature générale des Activités économiques dans les Communautés Européennes")

Section	Title
A -	Agriculture, forestry and fishing
B -	Mining and quarrying
C -	Manufacturing
D -	Electricity, gas, steam and air conditioning supply
E -	Water supply; sewerage, waste management and remediation activities
F -	Construction
G -	Wholesale and retail trade; repair of motor vehicles and motorcycles
H -	Transportation and storage
I -	Accommodation and food service activities
J -	Information and communication
K -	Financial and insurance activities
L -	Real estate activities
M -	Professional, scientific and technical activities
N -	Administrative and support service activities
O -	Public administration and defence; compulsory social security
P -	Education
Q -	Human health and social work activities
R -	Arts, entertainment and recreation
S -	Other service activities
T -	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U -	Activities of extraterritorial organisations and bodies

Source: Eurostat (2008)

Howells (2010) has even suggested that a "segmentalist" approach to innovation in services is emerging, reflecting the diversity of service activities and innovation patterns, and moving away from analysing services as a whole. Even a casual encounter with the literature on services is likely to reveal that for every generalisation that can be made about these activities, there will be numerous exceptions. (For example: Services only create intangible products—what then about this dental filling? Services cannot be stored—what about this website or computer software? Services are coproduced with their customers—what about this TV broadcast?, and so on.) But there are several features that are common to many services, even if many exceptions can be found. It is worth considering these commonalities. The foreground social and economic characteristics are quite distinctive from those typical of manufacturing, and the implications for innovation are significant.

There are many ways in which differences between manufacturing and services can be conceptualised—see Miles (1993) for an extensive list—but two interrelated features underpin most of these. The first of these features is the "intangibility" of the service product. While manufacturing is about making goods, service activities are about doing things—about changing (or reducing change in) the state of people, artefacts, symbols, etc. Intangibility is associated with such attributes of service products and processes as the difficulty in storing or transporting them, problems in patenting service innovations, and the difficulty in demonstrating the service in advance of purchase. This latter point accounts for the need for regulation of many services, and the challenge that can confront the service supplier when it comes to convincing consumers about the superiority of innovative services.

As noted, some services have quite tangible outputs. But typically the material costs of the dental filling or the optical disc is a small fraction of the cost of the professional labour involved in tailoring the filling or creating the information content of the disc. It is the less tangible aspects of the service that typically count as most important, and which can be difficult for prospective purchasers or clients to assess. One result of this is that many innovations from service organisations involve adding more tangible elements to the service (loyalty cards, for example), while others involve the creation of demonstrator products (demo discs, movie trailers, free trials) or certification of various forms (quality standards, membership of professional bodies, etc.).

The second key feature of services is their "interactivity"—referred to in other studies by such terms as "consumer-intensity" (Gartner and Reissman, 1974), and "servuction" (Eiglier and Langeard, 1987). This reflects the high level of contact, exchange, and "touch" that is involved in most services (Miles, 2005). It is helpful to think of the client as coproducing the service, not least because this implies that service innovation (if not necessarily innovation in service organisations) is liable to involve learning and behavioural change on the part of the user as well as the nominal service supplier.

The extent of interactivity can vary considerably. In the case of a consultancy service, there may be protracted discussion about just what the problem is, there may be in-depth questioning and observation from the consultant, the final report and recommendations may be presented in a variety of face-to-face settings, and need to be reflected upon and further explored by the client. In contrast, a bus ride may involve little more than turning up at the stop, buying a ticket, and sitting in a suitable seat until the destination is reached. Consultancy activities and bus trips can of course also vary immensely!

There are many forms that interactivity can take, for example:

- Since the interaction involves information exchanges, there is much scope for application of new Information Technology—from using powerpoints to support consultants' presentations and electronic whiteboards in teaching, to automated teller machines and online banking services. These new technologies are pervasive in service industries, and continue to be the site of considerable innovation, not least as the organisations learn new ways of enhancing their services through their application.
- Much innovation centres, too, on the distribution of activities between supplier and client, with "self-service" proving one popular approach—not simply because it can reduce labour inputs on the part of the service organisation, but also because it can improve the quality and efficiency of the experience for the customer. Such innovations require creation of a mutually acceptable framework for identifying and accessing the objects of the service, be this bank account details or consumables on supermarket shelves.
- The experience provided by many services is often dependent on the behaviour of multiple consumers, whether these are contributors to a social networking website, other passengers in public transport vehicles, or the other users of sports or cinema facilities. In some cases some sort of input from other users is needed to make the experience worthwhile, while in other cases the consumer may really prefer to be alone.

One consequence of interactivity is that the service supplier and client often need to be at the same place at the same time, though the use of Information Technology may reduce the need for this in respect of information services, at least. Other important consequences of this feature of many services are that service quality will be a matter not only of supplier effort, but also of the clients' own inputs; that productivity as measured by labour inputs by the supplier may be achieved at the cost of more labour required from the user; and that there is liable to be much heterogeneity among the outputs of a service organisation. Some services may be relatively standardised, but many others are customised or at least mass-customised by the assembly of the service out of multiple modules, put together according to customer requirements. Some other services are completely bespoke ones, specialised to a particular client requirement. Manufacturing varies as well, between mass production, mass customisation, and small-batch specialised production (rapid prototyping is a rare case of a service industry that actually makes physical goods—though the objective is to test the viability of designs). The heterogeneity of outputs contributes to the difficulty in assessing service quality prior to service production, and to the difficulties confronting service productivity measurement.

One of the main trajectories of service innovation has been what Levitt (1972) recognised four decades ago as the industrialization of services. As service firms grew, they could adopt a "production line approach," with more standardized products produced on an almost mass production basis, with high division of labour and use of technology. Increased standardization, we can now see, can accompany mass customisation, with standardised service modules combined in numerous ways to produce services whose quality varies little from branch to branch or franchise to franchise of hotels, fast food restaurants, supermarkets, and the like. Many firms in such sectors that adopt this model of service industrialisation are dependent on relatively high levels of low-wage and fairly unskilled staff, often working on part-time or insecure bases.

SERVICE INNOVATION AND NEW SERVICE DEVELOPMENT

This essay began by noting some ambiguities connected with the term "service", and in addition there is one sense of the term that is particularly important in terms of these common features. This is the sense that a service is something done for somebody—that a service is about providing value to another human being or set of human beings. (There are inevitable exceptions. In informatics it is commonplace to talk of computer systems and components as providing services to each other, as in Service-Oriented Architecture. Then there are services that are oriented to the well-being of natural environments, which may directly affect no human beings though they may be thought of as providing some satisfaction to those people who know about them.) This sense of service comes to the fore in the recent stream of work on "service-dominant logic" (see, for example, Lusch *et al.*, 2008, Vargo and Lusch, 2006). This stream originally stems from service marketing studies, though it has achieved much wider influence. It sought to move beyond approaches to service marketing that saw services as simply intangible products that could be dealt with by a little elaboration of the methods used for marketing goods. Instead, the focus is on service as a process: service is the process of applying resources to create benefits; and it is a coproduction process, where both "supplier" and "client" make contributions and gain benefits. All economic activity can be seen in this perspective as an exchange of services.

Of course, the amount and type of effort that is put in by the partners varies from service to service, but the point that service users are typically engaged in activities other than just purchase, and that these activities have a very important impact on the quality of the service that is produced and received, is important. This clearly relates to the notion of "interactivity" introduced earlier, and to research examining innovations involving these activities and the service relationship. Currently there is much interest in the ways in which service customers can be mobilised as "prosumers" to enhance each others' experiences, for example in Web2.0 and social networking applications, but there has been earlier work on how innovation may centre on the servuction process (e.g. Belleflamme *et al*, 1986).

Change in service relationships and associated experiences has also been a central theme in the emerging discipline of "service design". Recent years have seen many established industrial design firms move to handling aspects of service design, as well as the appearance of specialised firms with this focus. Education and scholarly research have only recently begun to catch up with this, with a journal of service design—*Touchpoint*—being launched in 2009. But a few early pioneers had written on topics such as service design and quality (Gummesson, 1990), so that sufficient material has been produced for reviews to be available (Moritz, 2005; Saco and Goncalves, 2008).[2] Among the things revealed in such accounts are the distinctive techniques required for service design, so as to reflect the co-evolution of user and supplier behaviours and experiences in the course of service delivery. Such techniques include service blueprinting, storyboarding, and interface and interaction design. The need for quite different design strategies from those that have been prevalent in industrial product design reflects the importance for services of such features as intangibility and interactivity.

Den Hertog (2000) makes the point that service innovators need to be attentive to technological opportunities, but should avoid a technologist's view of New Service Development (NSD) ("technologist" is used herein the Gallouj sense). They should rather consider what changes might be effected in addition in terms of *service concepts*, *client interfaces*, and *delivery systems*. This suggests that any innovation can be thought of as a combination of, and possibly as changes in, these dimensions as well as in the technologies being employed.

Much in the same way that discussions of service design and innovation have emerged rather recently, so the study of New Service Development (NSD) is relatively young—but rapidly growing, and already the subject of several reviews. In this case, Johne and Storey had already by 1998 been able to examine a range of studies, reflecting what we have earlier termed the interactivity of services. Customers and the understanding of their roles, expectations and experiences, are particularly important in NSD, given the likelihood that their cooperation is critical in shaping the quality of the service outcome. The employees who interact with customers also have to be taken into account—both as sources of insight and co-producers of the service; their informed cooperation is also vital. Typically studies of NSD attempt to identify which factors make for successful introduction of new services, with Martin and Horne (1993, 1995) also noting the need for customer and employee (and managerial) participation in the NSD process, together with strategic use of customer information. In the service firms they studied, specialized innovation functions were uncommon, and successful NSD was rarely achieved by a few experts. The NSD literature frequently pays attention to the strategies and characteristics of service-producing organisations. Considerable emphasis is placed on the role of service-employees training and broader learning opportunities, on the scope for sharing information and experiences, on the ease of establishing multifunctional project teams, and the like. Similar prescriptions also emerge from studies of manufacturing innovation, and it remains to be established whether successful NSD is really that different—and indeed, whether there are not huge differences across services of different types. But it is clear from numerous

"The need for quite different design strategies from those that have been prevalent in industrial product design reflects the importance for services of such features as intangibility and interactivity"

[2] See also the service design network, at http://www.service-design-network.org (accessed 24 August 2010).

studies (reviewed in Miles, 2005 2010, and elsewhere) that new services are rarely produced through formal R&D departments and/or production engineering—though such an approach is used in some very large service firms and technology-related services in fields such as information technology and engineering. More commonly, service innovation is organised through transitory project management structures—and much innovation emerges from ad-hoc, on-the-job experimentation. Surveys of innovative service firms (e.g. Arundel *et al.,* 2007, IOIR, 2003) suggest that—perhaps surprisingly—such firms tend to report less use of suppliers and customers as sources of information for innovation than do manufacturing firms. (In contrast, consultancies and competitors seem to be more important sources of information for service firms than for manufacturers.) One sector that does report more use of clients as sources of information is Business Services—where there is often a very deep level of interactivity. Wholesale and retail trade services are more likely to see suppliers as influential, as might be expected.

Sundbo and Gallouj (2000) suggest that several different ways of organising service innovation can be differentiated (their analysis can be applied to process innovation in service organisations as well as to NSD). Miles (2010, pp. 523-524) summarises their approach as indicating seven broad patterns, while noting that particular service innovations may be organized in different ways within the same organisation:

1. The classic R&D pattern, with specialized departments conducting research of a strategic nature does exist in some service organisations—mainly large and/or technology-based ones, as noted above.

2. The Services Professional Pattern is often found in knowledge-intensive organizations such as KIBS, whose professionals frequently generate solutions for clients that are ad hoc and highly customized. Their innovations typically rely on employees' professional skills. Much innovation intelligence may flow through professional networks and associations, or other communities of practice. Many consultancy firms, and some "creative industries" (e.g. advertising and design) follow such a model. One major challenge for these firms is "capturing" and replicating innovations that are made in practice by professionals, and much attention in knowledge management is directed to this.

3. A Neo-Industrial Pattern lies between patterns (2) and (3): alongside a specialized R&D or innovation department, there is much more distributed innovation in the course of professional practice. This often characterizes, for example, health services and some large consultancies.

4. The Organized Strategic Innovation Pattern is encountered in large service firms, such as airlines, hotel chains, and retailers. Innovation is organized in the form of projects that are directed by more or less transitory cross-functional teams, working through distinct steps of project management, and often with strong leadership from marketing groups.

5. An Entrepreneurial Pattern characterises start-up firms that offer services based on more or less radical innovations: these may be technological or rely more on new business models: many so-called gazelles, online services, and others follow this pattern, across many sectors: typically it is short-lived and they move into one of the other innovation modes.

6. The Artisanal Pattern is found in many smaller-scale and low-tech physical ("operational") services, such as clearing and catering. These are classic supplier-driven sectors, where major innovations are imported from other sectors (e.g. manufacturing), though innovation may also be driven by regulations and demand. Employees and managers may be sources of (typically incremental) innovation.

7. Finally, the Network Pattern involves a network of firms acting together, and adopting common standards or operating procedures. There may be a dominant company in such a network, and this has been the case in the rolling-out of such innovations as ecommerce, where often a major customer has requested that its suppliers use standardised means of electronic trading. Many services are organized in franchise networks through which such diffusion of innovations may take place: this is familiar in sectors such as fast food and hotels, and also in some professional sectors.

INNOVATION IN SERVICE INDUSTRIES

Overviews of innovation in service industries have been available for some years, too (e.g. Miles, 1994, and later reviews in 2005 and 2010), and much of the *Handbook of Innovation and Services* (Gallouj and Djellal, 2010) also considers this theme. These studies confirm the argument that the organisation of innovation in service organisations typically takes forms different from the R&D model supposedly characteristic of manufacturing. In fact, as already implied by Pavitt's (1984) taxonomy, many manufacturing firms do not follow this model—it is most common in high-tech firms and in larger firms in other manufacturing subsectors. (And, we might add, such firms do not always apply this model across their range of activities—the distribution and retail activities, and other product services, may evolve quite independently of the product innovation itself.)

Survey studies which allow for comparisons to be made across sectors have confirmed that service firms do introduce innovations, although overall the service sectors may have lower rates of such innovation than manufacturing firms overall. But there are high variations across different sections of services. The innovation budgets of service firms also tend overall to be lower than those of manufacturing firms, even when we compare firms of similar sizes (important because innovation behaviour tends to be strongly associated with firm size, and, as noted, most service sections are more dominated by small firms than is manufacturing). However, the various parts of the service sector differ considerably in terms of how frequently they innovate and how far they invest in innovation. While there are exceptions in all service subsectors, the general trend is for more physically-oriented services like transport and wholesale and retail trade to report lower levels of innovation, and for more information-oriented services, such as financial services and KIBS, for example, to be much more innovation–intensive. This result may be rather different from what might have been found had we undertaken such surveys in the 1920s, rather than the 2000s. In the first half of the twentieth century, the physical services were being transformed through the application of electrical energy and petroleum engines. By the turn of the century, it was the new information technologies that were being used to create new and improved services, and these were particularly important for activities such as financial and computing services, and all sorts of professional activity. Technology-related KIBS in particular—firms providing computer, and engineering services, for instance—typically have large innovation budgets.

The availability of large-scale surveys makes it possible to apply cluster analysis and similar approaches to identify and classify distinctive sets of firms or sectors. Thus Hipp and Grupp (2005) differentiated between knowledge-intensive, network-intensive, scale-intensive and external innovation-intensive patterns in German service firms. There were clear tendencies for certain types of service industry to follow particular types of innovation dynamic. The knowledge-intensive pattern, for instance, was particularly marked in technical and R&D services and computer services. The network-based model was prevalent in banking, while the supplier-dominated model was especially important in other financial services. But Hipp and Grupp also warn against a simple identification of sectors with innovation patterns. Though there are more or less strong trends, all sectors have their exceptions—and indeed, some cases of each of the innovation patterns were located in each of the service sectors.

These studies typically focus on issues such as innovation expenditure and the sources of information for innovation. Less attention is given to the nature of the innovations themselves, but several studies indicate that service firms are somewhat more prone than manufacturers to report non-technological and organisational innovations. Howells and Tether (2004) report that while a substantial share of service firms considered their main innovative activities to have been solely organizational, this was very rare among manufacturing firms. Kanerva *et al.* (2006) report that service firms (especially financial and wholesale sectors) are more prone to initiate organizational change; Schmidt and Rammer (2006) and Miles (2008) report that manufacturers and Information Technology service firms tend to emphasize technology-based innovation, while most services emphasize organizational innovation—though on the whole, sectors that are more technologically innovative sectors are also more organizationally innovative)

There are now numerous studies exploring the broad picture of service-sector innovation from CIS data (e.g. using CIS2 data for Europe, Tether *et al.* (2002; presenting CIS4 results, Arundel *et al.* (2007) and Eurostat (2008). Below, we shall focus on three particularly interesting services—KIBS, creative services, and public services.

KIBS

KIBS are generally classified into two groups—T-KIBS (technology-based ones) such as (computer services, architectural and engineering activities, technical testing and analysis, R&D services, etc—and P-KIBS, more traditional professional services such as legal and accountancy support, market research and management consultancy. Many studies have shown that KIBS overall tend to have high levels of innovation, and have suggested that they behave more like high-tech firms than other services. But Rodriguez and Camacho, (2010) analysed Spanish CIS4 data to show that there are actually quite different types of KIBS innovators. Some are like high-tech manufacturers—“hard innovators” who develop product innovations, largely based on internal R&D. But there were three other groups. “Knowledge diffusers” are those KIBS who act as agents of knowledge transfer, who have close relationships with other agents across the innovation system, including universities and public research institutes and technological centres. The other groups were "lonely innovators"—with few collaborations, reliant on their own innovation capacity for developing technological and/ or organisational innovations—and a small group of "soft innovators" that mainly develop organisational and process innovations, often largely based on acquisition of machinery and equipment. Clearly we need to be cautious in generalising about KIBS.

Equally clearly, the innovative activity of KIBS can be important for the whole economy. Even those KIBS whose primary role is not "knowledge diffusion" are active in providing solutions for their business clients' problems. Often these will involve helping the client to undertake innovations in practice or technology. Sometimes this involves a coproduction of innovation on both sides (den Hertog, 2000), as new knowledge is produced through the combination of the KIBS firm's generic understandings and the more local knowledge of the client. Through negotiation concerning the nature of the problem and the potential solutions, both service supplier and client can learn: the challenge is for these organisations to capture this learning and replicate the innovations.

Creative Services

In recent years, much policy attention at national and local levels has been focused on the "creative industries", which are mostly activities defined by their focus on end-user experience and production of creative content (though typically some sorts of entertainment—theme parks, sports—and cultural service—museums—are omitted from the classifications, while computer software is often included!). Some of the services here are aimed at businesses, and a category of C-KIBS might be added to the T- and P-KIBS, covering for example advertising, design, and graphics and multimedia services supplied to organisations. Until recently, such activities were seen as more the domain of media studies and cultural criticism than of innovation research, and there are certainly difficulties in trying to specify the innovativeness of a new fashion design or TV format, for example.

But we are now beginning to see studies focusing on innovation strategies in experience industries (e.g. Voss and Zomerdijk, 2007), and evidence is accumulating that demonstrates that creative industries engage both in familiar sorts of product and process innovation, and in many other forms of organisational and business model innovation (e.g. Miles and Green, 2007). With a few exceptions, these industries have been neglected in innovation surveys, despite being the focus of a good deal of policy rhetoric. Some of them play important roles in relating businesses to changing social milieux; some of them contribute to the creation of more innovative and creative milieux, as argued in many accounts of the creative city and economy.

Public Services

Finally, we briefly consider public services, which are at the centre of policy concern (with persistent concerns about their productivity and cost), and also have been neglected in innovation surveys (which almost always only examine private services). At a time of considerable reform of public services, and redrawing of the boundaries between public and private sectors, there is a striking absence of evidence on which to base policy.

It is widely argued that public sectors are less innovative than private firms (though the evidence on this is mixed—see Halvorsen *et al.*, 2005).[3] This is often seen to result from

[3] Online resources on public service innovation include the PUBLIN project (http://www.step.no/publin/), the *Innovation Journal* (http://www.innovation.cc/) and NESTA studies on public services (http://www.nesta.org.uk/assets/documents/ready_or_not and numerous other publications at NESTA's website). The ServPPIN project is an interesting recent effort to explore public-private innovation networks—see http://www.servppin.com/ (all accessed 20/07/2010).

lack of competition and bureaucratic (and political) control structures, so one very popular solution has been the reforms known as the "New Public Management", that introduce market-like structures and more entrepreneurial management into the public sector—with some moves towards public-private partnerships and privatisation (there is now a sizeable private "Public Services Industry" in some countries—Julius, 2008). Most public services consist of multiple "branches" of very large organisations, in many cases requiring highly skilled staff (doctors, teachers, etc.), though other cases involve more low-skill operational staff (cleaners, security staff, etc.). As large organisations, there is scope for economies of scale, and public sectors were early pioneers in the use of information technology for back-office functions. There is also scope to influence the innovation system more generally through public procurement, and "innovative procurement" has been a recent theme. However, the proliferation of local bodies and specialised professions, dealing with complex social issues, may create a risk-averse attitude to innovation, push it in inappropriate directions, or restrict the diffusion of innovations created in the course of practice. New Public Management alone is unlikely to resolve all of these problems, and thus we see numerous efforts to create new institutions that can identify and disseminate ideas for and examples of good practice and creative solutions.

CONCLUDING REMARKS

With service sectors being the bulk of the economy, and service forming an even larger share of all economic activities, it is difficult to present a succinct account of service innovation. What the studies reviewed here do point to is the need to explore innovation processes and trajectories that go well beyond those familiar from studies of automobiles, electronics and pharmaceuticals. They suggest, too, that we should be prepared to uncover a very wide range of different structures and strategies, which are evolving as the service economy continues to develop.

This has considerable implications for policy—one size will not fit all, and innovation policies will need to pay attention to the challenges of service innovation in a competitive world (as well as in public services). Likewise, new management capabilities, and training to support their development and deployment, are needed. Often the issue of cross-disciplinary and cross-professional team working rises to the fore, as innovations involve the combination of multiple goods and services in what has been dubbed a "product-service system", requiring knowledge of technologies, social institutions and regulations, and specific types of client and client interface.

Service innovation and innovation in services were themes that remained neglected for a surprisingly long time. Now they have risen to the fore, and are engaging the attention of researchers and practitioners of many kinds. One of the striking developments in the recent past has been the commitment of IBM and several other large firms—mostly but not exclusively those dealing with Information Technology services (and hoping to apply Information Technology within a huge range of service activities) to the creation of a new "science of service" or SSME (Services Science, Management and Education). This has been manifest in the foundation of a new journal of *Service Science*, the organisation of numerous conferences,[4] and several substantial publications (e.g. Maglio *et al.*, 2010) outlining new concepts

[4] See for example http://www.ibm.com/developerworks/spaces/ssme for links to conferences and online resources of many sorts, including curricula for new training programmes. A network of SSME researchers is organised through http://www.ssmenetuk.org/ and a symposium on "Succeeding through Service Innovation" is described at http://www.ifm.eng.cam.ac.uk/ssme/ (all sites accessed 31/10/2020).

of service systems and the various elements that might comprise a service science. The notion of a service science is a formidable challenge; even information-processing services take a wide range of forms, engaging suppliers and users in many ways. But the concentration of effort is already beginning to yield promising perspectives on the analysis and design of service systems, and we may well see the new thinking about service and services being reflected in new forms of, and strategies for, service innovation over the coming years.

REFERENCES

Arundel, A., M. Kanerva, A. van Cruysen and H. Hollanders (2007), *Innovation Statistics for the European Service Sector,* UNU-MERIT, INNO-Metrics Thematic Papers at: http://www.proinno-europe.eu/sites/default/files/page/10/07/Innovation_Indicators_for_the_European_Service_Sector.pdf Accessed 31/08/2010.

Avadikyan, A., and S. Lhuillery (2007), *Innovation, Organisational Change and Servicisation: a Micro Data-Level analysis in Five European Countries*, paper presented at DIME Workshop "Organisational Innovation: the Dynamics of Organisational Capabilities and Design", GREDEG—DEMOS, West-End Hotel, Nice, 15-16 November.

Barras, R. (1986), 'Towards a Theory of Innovation in Services', *Research Policy* 15(4), pp. 161-173.

Barras, R. (1990), 'Interactive Innovation in Financial and Business Services: the Vanguard of the Service Revolution', *Research Policy* 19, pp. 215-237.

Belleflamme, C., J. Houard, and B. Michaux (1986), *Innovation and Research and Development Process Analysis in Service Activities,* Brussels, EC, FAST, Occasional papers no. 116.

Coombs, R., and I. Miles (2000), "Innovation, Measurement and Services", in J. S. Metcalfe and I. Miles (eds.), *Innovation Systems in the Service Economy: Measurement and Case Study Analysis*, Boston, MA: Kluwer Academic, pp. 85-103.

den Hertog, P. (2000), "Knowledge Intensive Business Services as Co-Producers of Innovation", *International Journal of Innovation Management* 4(4), (December), pp 491-528.

Droege, H., D. Hildebrand and M. A. Heras Forcada (2009), "Innovation in Services: Present Findings, and Future Pathways", *Journal of Service Management* 20(2), pp. 131-155.

Eiglier, P., and E. Langeard (1987), *Servuction*, Paris: McGraw-Hill.

Eurostat (2008) *Science, Technology and Innovation in Europe,* Luxembourg: Office for Official Publications of the European Communities.

Fuchs, V. (1968), *The Service Economy,* New York: National Bureau of Economic Research.

Gadrey, J. (2002), "The Misuse of Productivity Concepts in Services: Lessons from a Comparison between France and the United States" in J. Gadrey and F. Gallouj (eds.) *Productivity, Innovation and Knowledge in Services: New Economic and Socio-Economic Approaches,* Cheltenham: Edward Elgar.

Gallouj, F. (1998), "Innovating in Reverse: Services and the Reverse Product Cycle", *European Journal of Innovation Management* 1(3), pp. 123-38.

Gallouj, F., and F. Djellal (eds.) (2010), *The Handbook of Innovation and Services*, Cheltenham: Edward Elgar.

Gallouj, F., and M. Savona (2010), "Towards a Theory of Innovation in Services" in Gallouj and Djellal *The Handbook of Innovation and Services*, Cheltenham: Edward Elgar.

Gartner, A., and F. Reissman (1974), *The Service Society and the New Consumer Vanguard*, New York: Harper and Row

Grönroosa, C., and K. Ojasalo (2004), "Service productivity: Towards a Conceptualization of the Transformation of Inputs into Economic Results in Services", *Journal of Business Research* 57, pp. 414– 423.

Gummesson, E. (1990), "Service Design", *Total Quality Management* 2(2), pp. 97-101.

Halvorsen, T., J. Hauknes, I. Miles and R. Røste (2005), *On the Differences between Public and Private Sector Innovation,* PUBLIN Project, report D3, Oslo: STEP; available at: http://www.step.no/publin/reports.html (accessed 31/08/2010).

Harzing, A. W. (2010), *Publish or Perish,* version 3.0.3883 (18 August 2010), available at http://www.harzing.com/pop.htm (accessed 20 August 2010).

Holmlid, S. (2007), *Interaction Design and Service Design: Expanding a Comparison of Design Disciplines,* Stockholm, Nordic Design Research—Design Inquiries available at http://www.nordes.org/data/uploads/papers/143.pdf (accessed 24 August 2010).

Howeils, J. (2010), "Services and Innovation and Service Innovation", in F. Gallouj, F. Djellal and C. Gallouj (eds.), *The Handbook of Innovation and Services,* Cheltenham: Edward Elgar.

Howells, J., and B. Tether (2004), *Innovation in Services: Issues at Stake and Trends,* Inno Studies Programme (ENTR-C/2001), Brussels: Commission of the European Communities available at: http://www.cst.gov.uk/cst/reports/files/knowledge-intensive-services/services-study.pdf (accessed 24 August 2010).

IoIR (Institute of Innovation Research) (2003), *Knowing How, Knowing Whom: A Study of the Links between the Knowledge-Intensive Services Sector and The Science Base,* Manchester, University of Manchester; Report to the Council for Science and Technology, available at: http://www.cst.gov.uk/cst/reports/files/knowledge-intensive-services/services-study.pdf (accessed 20 June 2009.

Johne, A. and C. Storey (1998), "New Service Development: a Review of Literature and Annotated Bibliography", *European Journal of Marketing* 32(3/4), pp. 184-251.

Julius, D. (2008), *Public Services Industry Review—Understanding the Public Services Industry: How Big, how Good, where Next?* London: BERR—Department for Business Enterprise & Regulatory Reform.

Kanerva, M., H. Hollanders and A. Arundel (2006), *Can We Measure and Compare Innovation in Services?* Luxembourg: European TrendChart on Innovation, 2006 TrendChart report available at http://archive.europe-innova.eu/servlet/Doc?cid=6373&lg=EN (accessed 31/08/2010).

Lusch, R. F., S. Vargo and G. Wessels (2008), "Toward a Conceptual Foundation for Service Science: Contributions from Service-Dominant Logic", *IBM Systems Journal* 47(1), pp. 5-14.

Maglio, P. P., C. A. Kieliszewski, and J. C. Spohrer, (eds.) (2010), *The Handbook of Service Science*, New York: Springer.

Martin, C. R., and D. A. Horne (1993), "Services Innovation: Successful Versus Unsuccessful Firms", *International Journal of Service Industry Management* 4(1), pp. 49-65.

Martin, C. R., and D. A. Horne (1995), 'Level of Success Inputs for Service Innovations in the Same Firm', *International Journal of Service Industry Management* 6(4), pp. 40-56.

Miles, I. (1993), "Services in the New Industrial Economy", *Futures* 25(6), pp. 653-672.

Miles, I. (1994), "Innovation in Services", in M. Dodgson and R. Rothwell (eds.) *Handbook of Industrial Innovation,* Aldershot: Edward Elgar.

Miles, I. (2005), "Innovation in Services", in J. Fagerberg, D. Mowery and R. Nelson (eds.), *The Oxford Handbook of Innovation*. Oxford: Oxford University Press.

Miles, I. (2008),"Pattern of Innovation in Service Industries", *IBM Systems Journal* 47, pp.115-128.

Miles, I. (2010), "Service Innovation", in P. P. Maglio, C. A. Kieliszewski, and J. C. Spohrer (eds.), *The Handbook of Service Science*, New York: Springer 2010.

Miles, I., and L. Green (2008), "Hidden Innovation in the Creative Industries", London: NESTA Research report HICI/13 at: http://www.nesta.org.uk/assets/Uploads/pdf/Research-Report/hidden_innovatiol_in_creative_Industries_report_NESTA.pdf (accessed 31/08/2010).

Moritz, S. (2005), *Service Design: Practical Access to an Evolving Field* Koln: Koln International School of Design, accessed from http://stefan-moritz.com/welcome/Service_Design_files/Practical%20Access%20to%20Service%20Design.pdf (accessed 24 August 2010)

Pavitt, K. (1984), "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory" *Research Policy* 13(6), pp. 343-373.

Pavitt, K. (1994), "Key Characteristics of Large Innovation Firms" in M. Dodgson and R. Rothwell (eds.), *The Handbook of Industrial Innovation,* Aldershot: Edward Elgar.

Rodriguez, M., and J. A. Camacho (2010) "Are Knowledge-Intensive Business Services so 'Hard' Innovators? Some Insights Using Spanish Microdata", *Journal of Innovation Economics* 5(2010/1), pp. 41-65. Available at http://www.cairn.info/revue-journal-of-innovation-economics-2010-1-page-41.htm (accessed 31/08/2010).

Saco, R. M., and A. P. Goncalves (2008), "Service Design: An Appraisal", *Design Management Review* 19(1), pp. 10-19.

Susman, G., A. Warren and M. Ding (2006), *Product and Service Innovation in Small and Medium-Sized Enterprises,* State College, Pennsylvania, Smeal College of Business; report to The National Institute of Standards and Technology (United States Department of Commerce) RFP 05-480-5824 Available at: http://www.smeal.psu.edu/cmtoc/research/nistnpd.pdf (accessed on 31/08/2010).

Tether, B., I. Miles, B. Blind, C. Hipp, N. de Liso and G. Cainelli (2002), *Innovation in the Service Sector: Analysis of Data Collected under the CIS2* Manchester: University of Manchester, CRIC Working paper no. 11, ISBN 1 8402 006X—redraft of 7,45 available at ftp://ftp.cordis.europa.eu/pub/eims/docs/eims_summary_98_184.pdf (accessed on 31/08/2010).

Vargo, S., and R. F. Lusch, (2006), "Service-Dominant Logic: What It Is, What It Is Not, What It Might Be", in R. F. Lusch and S. Vargo (eds.), *The Service-Dominant Logic of Marketing: Dialog, Debate, and Directions*, Armonk, New York: M. E. Sharpe.

Voss, C., and L. Zomerdijk (2007), "Innovation in Experiential Services—An Empirical View", in DTI (ed.), *Innovation in Services,* London: Department of Trade and Industry, pp. 97-134.

BBVA

Financial Innovation: A Balanced Look

Robert E. Litan[1]
Brookings Institution, Kauffman Foundation

Financial innovation—once an unquestioned positive for any economy—has been much less celebrated since the financial crisis that began in 2007 and the Great Recession that followed in its wake. Some leading economists, notably Paul Volcker (former Chairman of the Federal Reserve), Paul Krugman (Nobel prize-winning economist at Princeton and a *New York Times* columnist) and Simon Johnson (former chief economist of the IMF), have each expressed skepticism about the social value of financial innovation in general, and with much justification, since some recent innovations helped lead to the crisis. More significantly, the sweeping Dodd-Frank Wall Street Reform and Consumer Protection Act enacted in the United States in summer of 2010 for the purpose of preventing financial crises or at least minimizing their harmful impacts, contains numerous provisions that, depending on how they are implemented by future regulators, could slow future financial innovation.

My thesis here is that financial innovation does not deserve all the blame that has been heaped on it. In fact, there have been a number of "good" innovations over the past several decades, though the critics are right that there have been "bad" ones as well. In this chapter I attempt to sort out the two, concentrating primarily on innovations introduced into the US market since the 1960s, the period in which Chairman Volcker, in particular, claims there has been little socially useful innovation. I conclude by offering some suggestions on how policy makers and regulators can best facilitate future "good" financial innovations while weeding out the "bad" ones before they can do much damage.

FINANCIAL VS. REAL SECTOR INNOVATIONS: ARE THEY DIFFERENT?

Before I turn to the main topics of the chapter, however, it is important to address a threshold issue that lurks behind the critique of financial innovation: that it is somehow different from innovations in the real economy, and therefore deserving of more skepticism. Several such differences have been alleged. Do the allegations have some basis?

One common view, for example, is that financial innovation consists of little more than arbitrage, and that even successful

[1] Senior Fellow, Economic Studies Program, The Brookings Institution and Vice President, Research and Policy, The Kauffman Foundation. Excellent research assistance by Adriane Fresh is greatly appreciated. This chapter is based on an earlier, longer essay, "In Defense of Much, But Not All, Financial Innovation," available on the website of the Brookings Institution, at www.brookings.edu.

trading strategies provide no net social value, since for every winner there is a loser. This view is misplaced. By definition, arbitrage eliminates differences in prices of like products or assets in different locations. Innovations that permit this to be accomplished more quickly and efficiently reduce costs, provide more accurate price signals at any point in time, and generally make markets more liquid and efficient. Added liquidity and efficiency, in turn, make it easier and less costly for firms to raise capital.

To be sure, there is something to the "zero-sum" critique of trading innovations. But as we will soon see, many useful financial innovations over the past several decades have little or nothing to do with trading.

A second very real difference between innovations in finance and those in the real sector is that the former are often more heavily leveraged, or financed with debt. Among many other things, the recent financial crisis has illustrated how dangerous excessive leverage can be. For example, the infamous collateralized debt obligation (CDO) that enabled far too many subprime mortgages to be originated and sold (with the mistaken blessings of the credit rating agencies) and which fueled the wider real estate bubble of the last decade, was a debt instrument that turned out to have largely destructive consequences. Likewise, the "structured investment vehicle," or SIV, was one of the primary means by which large banks financed CDOs. I have more to say about both these innovations below.

By contrast, real sector innovation tends to be financed more by equity than debt. The "dot.coms" of the late 1990s, many of which disappeared rapidly when the stock market bubble popped in April 2000, got their start largely with equity provided by angel investors or venture capitalists, and later were financed through public stock sales. Although the stock market implosion caused significant losses to stockholders—and not just those with investments in the dot.coms—these losses for the most part were not compounded by leverage. As a result, the real sector fallout from the dot.com bust was far less damaging than the damage caused by the popping of the real estate bubble.

The third alleged difference between financial and real sector innovations is that the former are often said to be driven largely by the desire to circumvent rules and statutes, while the latter are supposedly motivated overwhelmingly or even entirely by innovators seeking to do or make something "faster, cheaper, better." This distinction is not as clear as some may believe. Both financiers and manufacturers and service providers engage in games of "cat and mouse" with their regulators. Moreover, these games are not necessarily socially pernicious. To the contrary, efforts to get around "bad" or inefficient rules—such as the Depression-era limits on interest that banks could pay their depositors—are to be applauded. Other efforts (such as the creation of the supposedly off-balance sheet SIVs by banks) that get around "good" rules, such as capital standards, deserve condemnation.

My bottom line: when financial innovation leads to a better, faster or cheaper outcome, it is no different from real sector innovation. It is only when financial innovation is defined by or augmented by leverage that it can significantly differ from and be more dangerous than real sector innovation.

WHY DOES FINANCIAL INNOVATION MATTER?

Financial institutions, markets and instruments perform four broad social and economic functions. Innovations that improve

> "When financial innovation leads to a better, faster or cheaper outcome, it is no different from real sector innovation. It is only when financial innovation is defined by or augmented by leverage that it can significantly differ from and be more dangerous than real sector innovation"

the way these functions are carried out, by definition, are useful. Others may temporarily appear to be improvements, but in fact turn out to have socially pernicious collateral effects. Before I give specific examples of both types, it is important to know what these functions are.

The first function of finance is to provide a *means of payment and storing wealth.* What we know as "money" was first embodied in coins, livestock and foodstuffs, later in paper, and then in bank checking accounts. More recently, "money" has become digitized, found on general purpose credit cards (first introduced by American Express and Bank of America in 1958), and is transferred in large wholesale amounts electronically, through automated clearinghouses, the Fedwire system operated by the Federal Reserve (which also plays a central role in clearing checks written on commercial banks), and among major banks in the United States and other developed economies through the Clearinghouse for Interbank Payments System (CHIPS).

Second, by offering methods for earning interest, dividends and capital gains on monies invested in them, financial institutions and instruments *encourage saving*. Saving is socially important because it funds investments in physical and human capital that in turn generate higher incomes in the future. By the 1960s, most of the major institutions that exist today for facilitating saving were already in place: banks (and savings and loans, or specialized banks placing deposits in residential mortgages), insurance companies, mutual funds (but not yet money market funds or the increasingly popular index funds, each discussed below), and corporate pension plans. In addition, investors who had the time and the means could buy various financial instruments directly: government or corporate bonds and corporate stocks.

Third, financial institutions and markets, if they are working properly, *channel savings, whether by domestic or foreign residents, into productive investments.* Well before the 1960s, it was commonplace for companies wanting to build new buildings or to purchase new capital equipment, to borrow the funds from banks or insurers (whose primary functions as "financial intermediaries" are to direct the funds placed with them into private and public sector investments), or to issue new bonds or even to sell additional stock. The US government was instrumental in facilitating investment in residential housing by aiding the mortgage market. In the 1930s, the government established the Federal Housing Administration to insure mortgages taken out by low and moderate income households and the Federal Home

Loan Bank System. Later and over time, the federal government launched Ginnie Mae, Fannie Mae and Freddie Mac to provide a secondary market for most mortgages. Likewise, the government facilitated investment in education, or "human capital" by guaranteeing loans for post-secondary education. And, of course, the US government funded some of its own investment activities—such as the construction of the interstate highway system and the support of much scientific research—in part by issuing bonds (when tax revenues were not sufficient to pay for all this). But with the exception of a few small "venture capital" limited partnerships, to be discussed shortly, the US financial system had yet to develop by the end of 1960s a reliable institutional way to fund the inherently risky process of firm formation and initial growth.

A final, and sometimes overlooked, essential function of finance is *to allocate risks to those who are most willing and able to bear them.* This function is sometimes confused with the belief that finance reduces overall risk. Clearly it does not, and cannot. Instead, the best that finance can do is to shift risk to those who most efficiently can bear it and to spread it out so that it is not unduly concentrated among a handful of parties. One of the most important lessons from the financial crisis is that the kinds of "securitization" that made subprime lending possible did not end up de-concentrating risks of default of the underlying mortgages, as so many market participants and other analysts (including me) argued or expected.

As of the 1960s, insurance companies dominated the risk-bearing/allocation function of finance by underwriting various personal (auto, house, life, and health) and commercial (primarily property-related) risks. But insurance for *financial* risks was limited. Insurers were willing to bear the risk of default of municipal and state bonds and certain other financial instruments, but generally no other financial risks. Futures contracts for various commodities, which had long been present, were available on futures exchanges, but there were no widely-used mechanisms yet for insuring against other kinds of financial risks—due to fluctuations in interest rates, currencies (there was no need for this as exchange rates between currencies were fixed under the post-war Bretton Woods agreement), or stock prices (of individual companies or indexes). As will be discussed in the next section, various financial innovations over the past several decades have filled this gap.

ASSESSING RECENT FINANCIAL INNOVATIONS

It is now time to list and assess specific financial innovations since the 1960s, or around the time that the ATM machine, the one financial innovation celebrated by Chairman Volcker, was introduced into the US banking system. I assess the innovations qualitatively along three dimensions: the degree to which they widen access to a particular financial service, enhance convenience of users, and add to or detract from GDP or productivity. For each dimension, I give scores ranging from -- to ++ on what I believe to be the net impacts. Table 1 summarizes the results of my assessments, which I now describe verbally with respect to the innovations relating to each of the four financial functions just identified.

Innovations in Payments

The ATM machine is not the only socially useful innovation in payments in recent decades. As noted, general purpose credit (and debit) cards—both introduced since the 1960s—have become key parts of the

Table 1. Scoring Net Impacts of Recent Financial Innovations: A Summary

	Access	Convenience	Productivity/GDP
Payments			
ATMs	++	++	+
Credit card expansion	++	++	+
Debit cards	++	++	+
Saving			
Money market funds	++	++	0
Indexed mutual funds	++	++	+
Exchange-traded funds +	+	0/+	
Limited partnerships			
Hedge Funds	0	0	0/+
Private Equity	0	0	+
TIPS	++	++	0/+
Investment			
Credit scoring	++	++	0
ARMs	++	N/A+	-/--
Home Equity Lines of Credit	++	++	-
Asset-backed securities	++	++	-/+ (see text)
CDOs*	++	++	--
SIVs*	++	++	--
Rise of Venture Capital	+	+	++ (but future not clear)
Risk-Bearing			
Options/Futures Exchanges and Pricing	+	+	+/++
Interest/Currency Swaps	++	++	+/++
Credit default swaps	+	+	+
Permutas de cobertura por incumplimiento crediticio	+	+	+

*The positive scores here were temporary

Source: Analysis in text.

payments systems in developed economies, and recently, some emerging markets (notably China). On balance, despite some continuing complaints about credit card disclosures and some pricing practices, these “plastic monies” have been welfare-enhancing. Credit and debit cards have not only become more convenient and in some respects safer than money (in the United States, consumers are liable for only $50 for fraudulent uses or stolen cards, whereas they can lose all the money they carry if they are robbed). Credit cards also permit users to borrow, which has widened consumers’ access to unsecured credit, at lower cost, than was previously available, if at all, in grey markets, pawn shops or loan sharks. In the United States in particular, credit card borrowing also has enabled many entrepreneurs to launch their businesses at a scale beyond what would be possible through the entrepreneurs’ liquid net worth alone. Perhaps more for this reason than any other, credit cards in this fashion have made some positive contribution toward long-run economic growth.

Savings Innovations

Up until the 1970s, American retail investors (and those outside the United States) had relatively limited choices about where to put their savings: in bank deposits or their functional equivalents, in bonds, stocks of individual public companies, and a limited range of mutual funds. A series of financial innovations since then, driven in part by interest rate deregulation and in part from the commercial application of academic insights, have greatly expanded the range of available options.

The new choices include: money market mutual funds (developed as a way of circumventing interest- rate controls on bank deposits, a clear example of a "bad" regulation), indexed mutual funds (originating from the academic insight that actively managed funds rarely outperform the indexes), exchange traded funds or "ETFs" (a cost-saving innovation developed by the financial industry), financial limited partnerships, such as hedge funds and private equity funds ("alternative" assets that until relatively recently outperformed more liquid stock portfolios, though often because they are leveraged), and inflation-protected government bonds or "TIPs" (promoted by academic economists, first adopted in the United Kingdom in the 1980s and roughly ten years later, by the United States).

Broadly speaking, these savings vehicles have expanded access and convenience for investors, while most likely modestly enhancing economy-wide GDP. All of this occurred even while, in the United States, the private savings rate itself declined to roughly 0, most likely because households counted on the rising values of their homes to add to their wealth, a belief that had some validity until the real estate bubble popped in 2006-2007. Since the recession and the steep fall in both real estate and stock prices that followed, private savings rates have rebounded somewhat. Through mid-2010, when this chapter was written, the broader range of savings choices does not appear to have given much comfort to many newly risk-averse households, perhaps with the exception of TIPs and some commodity ETFs. These two instruments have come to be viewed as hedges as against the possible future inflation that may eventually follow in the wake of the significant monetary easing engineered by central bankers in the 2008-09 period to keep developed economies from falling even deeper into the recession triggered by the financial crisis.

> "For decades since the Depression, the US government has adopted a variety of measures to promote home ownership, including the creation of agencies to insure, purchase, and guarantee securities backed by residential mortgages"

Intermediation-Related Innovations

Ask most economists about the function of finance they believe to be most important, and the answer they will most likely give is the effective translation of savings into socially productive investments. Yet it is the innovations in financial "intermediation" that have proved to have had the most mixed record of all of the innovations surveyed in this chapter, and largely for this reason, these innovations deserve more discussion than the others.

In retrospect, the mixed record of the intermediation-related innovations is due overwhelmingly to what turned out to be hugely costly innovations in housing finance in particular. The outcome was not an accident, however, as it was heavily influenced if not directly promoted by US government policies that took the advancement of home ownership too far and by regulatory policies and attitudes that failed to police obviously unproductive innovations. To whom should blame then be assigned: the unwelcome innovations or the policies that facilitated if not directly led to them? The answer, of course, is both.

The general story of how all this happened is by now, of course, well known. For decades since the Depression, the US government has adopted a variety of measures to promote home ownership, including the creation of agencies to insure, purchase, and guarantee securities backed by residential mortgages. In addition, the federal tax code has long permitted taxpayers to deduct the interest paid on home mortgages. In 1978, the federal government encouraged depository lenders to extend credit to lower-income households and to other borrowers in low-income neighborhoods.

For years, this combination of incentives and mandates steadily lifted the home ownership rate to roughly 64-65% of all households by the mid-1990s. Both the Clinton and Bush Administrations, as well as Congress, wanted that rate to go still higher, premised largely on the notion that home ownership yielded important externalities: homeowners tend to take greater care of their homes and have more interest in the welfare of their neighborhoods than renters. In addition, the Bush Administration in particular viewed home ownership as one, albeit very important, element of its larger interest in developing a broad "ownership society." On this view, the greater the ownership interest people have in a broad array of assets—homes, companies, and the like—the greater will be their affinity for market-oriented policies.

Whatever the precise reason or mix of reasons, the strong bipartisan consensus favoring increased home ownership required that mortgage credit be made available on affordable terms to individuals and families with lower and less stable incomes than those who had earlier borrowed money to buy a home. This outcome, in turn, could only be realized if mortgage underwriting standards and down payments were relaxed for these "subprime borrowers."

The financial industry responded with a series of innovations, encouraged by federal policy, but also only made possible by a record bubble in housing prices to which both the innovations and the policies they fostered contributed. In combination, these innovations gave millions of Americans with less-than-stellar credit histories, many (but not all) with low incomes, access to mortgage credit, eventually pushing the home ownership rate to 69% of all households. This feat could not have been accomplished unless each and every one of the innovations about to be described had been developed and successfully marketed. In fact, some of these innovations were innocuous or mildly constructive on their own but in combination, they proved to be deadly dangerous to the financial system and ultimately to the rest of the economy:

- To give subprime borrowers access to seemingly "affordable" mortgages, the mortgage lending industry (banks to a limited extent, but most importantly, a new class of mortgage lenders that were not adequately supervised or regulated by federal authorities) invented a new variation of the adjustable rate mortgage which charged borrowers very low initial "teaser" interest rates that reset to rates substantially higher than the Treasury benchmark rate several years later when presumably higher real estate prices would permit the borrowers to easily refinance. In fact, this Ponzi-like scheme worked, up until about 2006, when home prices quit rising, which triggered higher subprime mortgage defaults, and eventually a full-blown financial crisis.
- Mortgage lenders would not have been comfortable extending subprime loans without the invention and later propagation by both major commercial and investment banks of a new financial instrument—the collateralized debt obligation (CDO). The CDO transformed what once was a socially productive innovation, a "mortgage-backed security" (MBS) backed by a pool of mortgages extended to prime quality borrowers, and turned it into a financial Frankenstein. The CDO's chief "innovations" included the use of newer securities backed by subprime (instead of prime) mortgages (increasingly underwritten without verification of the borrowers' income or job history) and the slicing of the cash flows from those securities into different classes or "tranches" (as they came to be called) designed to appeal to investors with different appetites for risk. Those investors wanting the safest instruments got first rights to the cash flows from the mortgages, while investors in less safe, but higher-yielding tranches had later claims. CDOs were central to the over-development of the subprime mortgage market because they allowed the originators of the mortgages to offload them (and thus not care about their quality) to the buyers of the different tranches of the securities.
- CDOs, and specifically the first and supposedly their safest tranches, could not have been sold, however, without other parties and still other innovations. Even with their first call on the cash flows, the first tranche of the CDOs would not have been attractive to risk-averse, yield-hungry investors (who were starved for safer, higher yielding securities in the low-interest environment engineered by the Federal Reserve to sustain the recovery) unless the ratings agencies bestowed their coveted AAA ratings on these particular securities (the lower ratings for the riskier tranches were less important to the other investors with higher risk tolerances). When the CDOs were first developed, the ratings agencies balked at doing this—the right instinct because the mortgages, after all, *were subprime*, and the agencies had no actuarial data of how these mortgages fared over an entire business cycle—but were ultimately persuaded when the banks and their allies made use of yet another recent innovation, the *credit default swap.* Although it has been much maligned in the popular press and in some journalistic accounts of the financial crisis, there is nothing inherently evil about the CDS itself: it is, after all, the functional equivalent of insurance and it is precisely for this reason that when packagers of CDOs added CDS protection, or even explicit bond insurance, to the

first tranches, that the ratings agencies then awarded their AAA ratings. The CDS later became infamous, not because of its flawed design, but because one of its largest issuers, AIG, failed to provide sufficient collateral when it was due under the terms of the contract. But not all CDO tranches (or even all CDOs) had protection from CDS or bond insurance, and thus investors in them lost heavily when subprime mortgages began defaulting after residential real estate prices quit climbing and started to plummet. Ultimately, therefore, by facilitating the unbundling of mortgage origination from the risk of holding the mortgages to maturity, CDOs greatly weakened, if not destroyed, lenders' incentives to underwrite mortgages prudently.

- One other dangerous financial innovation helped make the subprime mortgage debacle possible: the creation and greatly expanded use of the "structured investment vehicle" by some of the largest commercial banks active in creating and marketing CDOs and other risky asset-backed securities. SIVs were an entirely legal, ostensibly off-balance sheet, way for the banks to park their CDOs for sale to the public without having to raise or hold additional bank capital, as was required by the prevailing bank capital rules. The SIVs had an Achilles heel, however: they were funded almost entirely (except for a small sliver of capital supplied by the banks and outside investors) by short-term commercial paper, which although it was secured by the CDOs and other assets, proved to be highly susceptible to a creditor "run". In fact, this is just what happened in mid-2007: when the market value of this collateral began to fall as home prices turned down, buyers of this "asset-backed commercial paper" refused to roll it over or buy new issues. This triggered a wider run on many bank-sponsored SIVs, and after an aborted attempt by the Treasury to organize an industry-financed bailout of all major SIVs, the banks wound down their SIVs and assumed their assets, and more importantly their liabilities.

In sum, SIVs proved to be temporary, but ultimately flawed, vehicles for financing CDOs before they could be off-loaded to third-party investors. In this activity, they were joined by CDOs, teaser rate ARMs, and other mortgage-financing innovations that for a time provided access to mortgage credit to a wider pool of homebuyers with sub-par credit histories. In combination, these innovations fueled the bubble in real estate prices generally that eventually popped and brought the whole subprime mortgage experiment to an abrupt and highly painful halt.

Intermediation-related innovations, including some affecting mortgage lending, have not all been socially destructive, however, and thus this review would not be complete without mentioning the few that have made financial intermediation more efficient and improved social welfare during the past several decades.

A good place to start is with the securitization process, which began with mortgages in the early 1970s and later spread to other assets that served as collateral. The fundamental idea behind securitization—standardizing loans and using them to back securities to attract financing from the capital markets and not simply from banks—was and remains sound. A broader pool of financing, even without an implicit government guarantee, leads to modestly lower interest rates and thereby facilitates investment. But the subprime mortgage debacle revealed a significant downside of this new "originate-to-distribute" lending model. The separation of the origination of loans from those who ultimately hold them undermines incentives of the originators to be prudent. To some degree, securitization contracts address this problem by giving those who buy the loans rights to "put" them back to the originators under certain conditions. But these conditions are narrow and often contested. The better answer is for originators and securitizers to retain some credit risk—the Dodd-Frank bill requires 5% except for securities meeting high underwriting standards—so that both have stronger incentives for careful underwriting.

A second, related financial innovation is the development and now widespread use of credit scoring algorithms for personal and business borrowers. Credit scores have improved lenders' ability to predict and thus to better price risk through the interest rates they charge on loans. Credit scoring has widened availability of credit and, by making credit ratings more objective, has reduced (though not entirely eliminated) racial discrimination by lenders.

Finally, there has probably not been an iconic financial innovation over the past four decades in the US that has more greatly improved the allocation of savings toward productive investment than the formalization and subsequent rise of the venture capital industry, and to a lesser extent, of angel investing (equity injections in startups by wealthy individuals or groups of them). Venture capital firms—limited partnerships managed by the venture capital general partner(s)—have been rightly credited with having given birth to some of America's most famous companies, including Google, eBay, Amazon and Genentech, among others. Over the past decade, especially since the bursting of the Internet stock bubble, venture capitalists have moved away from "seed" investing—providing the initial equity to help launch companies—toward less risky later financing "rounds." Not surprisingly, returns for limited partners in VC have plummeted. Accordingly, after its remarkable successes of the 1980s and 1990s, the VC industry is now at a crossroads. Not only do few VCs provide seed capital of any kind, but the industry is wary of financing capital-intensive startup firms, such as those attempting to develop new drug therapies or "clean energy" alternatives to carbon fuels. Seed financing arrangements therefore are a ripe area for future innovation.

Financial Innovations and Risk-Bearing

The fourth key function of finance is to spread or allocate risk to those parties most willing and able to bear it. In recent decades, a variety of derivatives—financial instruments whose value depends on some other "underlying" asset—have proliferated to take on this role: exchange-traded stock options and financial futures and "over-the-counter" swap arrangements (relating to exchanges of cash flows with different interest rates; in different currencies; and to possible loan defaults, or "credit default swaps").

Derivatives have also become a media poster child for what supposedly went wrong in the financial crisis. Mostly, this is because one of the largest sellers of mortgage credit derivatives in particular, AIG, was brought to its knees in the fall of 2008 and was essentially taken over by the federal government (through a massive bailout orchestrated by the Federal Reserve) in an effort to contain systemic risk when that risk was at its peak. Mortgage-related credit default swaps also seemingly played the villain in Michael Lewis's highly popular and well-researched book, *The Big Short,* which contributed to the popular ire over this particular financial innovation.

Several points should be made about financial derivatives to put them in their proper perspective. First, financial derivatives are hardly new: options (the right to buy a product or instrument at a fixed price by a certain date) and futures (which require the holder to buy or sell the product or instrument at a fixed price at the maturity date) are several centuries old, and neither played any role in the recent financial crisis. Second, the vast majority of the hundreds of trillions of dollars (in nominal or face value) of the newer financial derivatives—those whose value is tied to movements in interest rates and currencies—also played no role in the crisis. Furthermore, these derivatives have been socially constructive: both interest rate and currency swaps permit parties with different risk preferences to act on them, without having to sell the underlying instruments (loans or bonds) to which the swaps refer.

Third, even the credit default swap, perhaps the most maligned of all financial derivatives, is fundamentally a constructive innovation. As a device to insure against loan default, the CDS affords a way for various parties—lenders, suppliers, customers, and others—to hedge against a very specific adverse event. Even when these swaps are bought by "speculators," or those without an economic interest in the underlying debt, they serve a useful function, as long as those selling these instruments have posted adequate collateral or margin and have sufficient capital to honor the contracts (as AIG did not). For one thing, without speculators, as in options markets where buyers typically do not own the underlying security on which the option is based,

> "In all markets there are winners and losers, but the fact that there are both does not condemn markets as institutions. Financial derivatives are no exception"

hedgers would have much more difficulty finding counterparties. Perhaps just as important, because CDS markets are typically far more liquid than are the markets in the underlying loans or bonds, CDS prices—which reflect the views of both hedgers and speculators—provide more accurate, timely market-based signals about the financial health of the company (or other issuer) whose debt is subject to the swap, than the market prices of the debt itself.

Fourth, the mortgage-related CDSs that have attracted so much attention and notoriety account for less than 5% of the overall debt market. Virtually all other CDSs relate to corporate bonds or loans.

Still, over-the-counter (OTC) markets for financial derivatives are far from perfect. They are dominated by a handful of dealers, prices are less than transparent and not timely, and as the AIG episode demonstrated, they are subject to breakdown if one or more large participants cannot honor their obligations. These problems should be largely, if not entirely, addressed once all the regulations relating to OTC-traded financial derivatives in the Dodd-Frank financial reform bill are developed and implemented.

In particular, the bill's requirement that standardized OTC financial derivatives of all types must be cleared by a central clearinghouse should remove the risks of cascading defaults entailed when derivatives contracts are bilateral and the counterparties can look only to each other for performance—and not to a central organization with the ability to set and enforce margin and collateral requirements. In addition, the bill requires such standardized contracts to be traded on quasi-exchanges ("swap execution facilities", a term yet to be defined by regulators) and for their prices to be reported more frequently than is currently the case. Better transparency, in turn, will make it easier to specify appropriate margin requirements and thus further reduce—beyond central clearing itself—the potential for system-wide collapse if one or more key derivatives participants are unable to honor their obligations.

In sum, financial derivatives enhance the ability of both financial and non-financial parties in an economy to hedge and control their financial risks. That some parties may use these contracts to bet correctly on the failure of particular companies or entire markets—as was the case with subprime mortgages—does not disprove this central proposition. In all markets there are winners and losers, but the fact that there are both does not condemn markets as institutions. Financial derivatives are no exception.

However, recent events have underscored the possibility that precisely because derivatives have become such important financial instruments, and that some of the parties who trade them are heavily interconnected with each other and with the rest of the financial system, it is essential to have an appropriate infrastructure to ensure that performance difficulties of one or more parties do not spill over and threaten to destroy the viability of the entire system. Recent legislative reforms in the United States, likely to be emulated in other countries, should substantially mitigate this risk.

PUBLIC POLICY TOWARD FINANCIAL INNOVATION IN THE FUTURE

The US Congress enacted and the President signed into law in the summer of 2010 the Dodd-Frank bill, the most sweeping financial reform legislation enacted since the Depression. This is not surprising. The financial crisis of 2007-2009 was also the most traumatic such event since the 1930s, and it would have been shocking had Congress done nothing in response.

Debate will surely continue for years over the merits of Dodd-Frank, as it was enacted and more importantly, as it will be implemented through the more than 200 rulemakings that will eventually convert its generally ambiguous statutory language into more concrete regulatory guidance. As regulators go about their jobs, and as future legislators consider tweaks or possibly major changes to Dodd-Frank, their attitudes toward financial innovation will be critical.

For example, if a skeptical view of financial innovation takes hold—either because the benefits of innovation are perceived to be presumptively small and/or the risks of catastrophic damage are feared to be non-trivial—then policymakers (and even voters) are likely to demand some sort of pre-emptive screening and possibly design mandates before financial innovations are permitted to be sold in the marketplace. This attitude would put regulators rather than the market in charge of screening innovation, a process that runs significant

"So, which model should apply to financial innovation in the future: the preemptive drug/nuclear power approach, or the wait-and-see policy generally pursued in most other contexts?"

risks of chilling innovations before they have even had the chance to be tested in the marketplace. Conversely, a more open, wait-and-see approach to innovation would wait for the innovations to emerge and then only regulate them if they generated costs greater than their benefits. This has been the reining-in approach to financial innovation so far, and it clearly is the way US policy has generally handled innovation in the real sector of the economy. A wait-and-see policy gives the market the first crack at screening innovations, but runs the opposite risk from the preemptive approach: if regulators are late to act, because of their own laxity or due to strong political counter-pressures, they can permit socially harmful innovations to wreak considerable havoc before they are reined in.

In some areas of life, of course, it is appropriate for policymakers to take a skeptical approach toward innovation. The concern about possible catastrophic outcomes is the reason Congress established the Food and Drug Administration, requiring among other things, that new drugs be tested extensively in both animals and humans, before they can be sold to consumers. Analogously, the dangers of a core meltdown of a nuclear reactor, however remote, have driven policymakers from the very beginning of the nuclear age to require the utilities that construct such facilities to comply with specific design and performance standards. The European Commission has gone further by adopting the "precautionary principle" in a number of arenas—environmental policy, for food products, and consumer protection generally. Although this principle has been applied differently in different contexts, it essentially means that where there are plausible grounds for believing that a new (or currently-existing) product or activity poses a risk to human health or the environment, policymakers can regulate it in advance (or even ban it).

With the narrow exceptions of pharmaceuticals and nuclear power, however, US regulatory and social policy has not followed the precautionary principle, and thus has so far taken a very different course from that in Europe. Government regulates only once evidence of detrimental side-effects becomes reasonably clear, and then, where the underlying statute permits, only when the benefits of regulating outweigh the costs, and the content of the regulation represents the least costly way to achieve those benefits.[2]

So, which model should apply to financial innovation in the future: the preemptive drug/nuclear power approach, or the wait-and-see policy generally pursued in most other contexts? Regulators in the United States and elsewhere will be wrestling with this question in many contexts for years to come. Despite the clear damage caused by this most recent crisis, I believe that, generally speaking, financial innovation should continue to be

[2] The balancing of benefits and costs and the least-cost requirement for regulation have been embodied in one fashion or another in Executive Orders since the Ford administration.

screened by the market first and regulators later, for at least two reasons.

First, unlike in the pharmaceutical industry, it is difficult if not impossible to conduct clinical trials in the financial sector. New drugs can be and are tested on sample populations, first for their safety and then their efficacy. If they pass both tests, the FDA makes the reasonable assumption that the behavior of the drug among a representative population may be extrapolated to the larger population. In contrast, while one theoretically could test a new financial product—say, a mortgage, on a sample population—its behavior among that sample is likely to be heavily time-dependent. Subprime, adjustable-rate mortgages with low initial "teaser" interest rates that were extended in 2002 and 2003 may have been quite safe because real estate prices then were rising, enabling borrowers to refinance at a later point. But the very same mortgages extended in 2006 or 2007, just when real estate prices had peaked, would and did have a different delinquency record. In short, the results of a "clinical test" of a financial product at a particular point in time cannot be safely extrapolated for *all time*.

One possible response to this problem, of course, would be to test new financial products over an entire business cycle before letting them on the market. But this would subject such products to substantial delay, and thus surely cut down on incentives of financial institutions to innovate.

Second, the danger that preemptive screening will chill productive innovation, even without the kind of regulatory delay just suggested, is another powerful reason to generally reject the preemptive approach to regulation in the financial arena. Thus, had US policymakers followed the precautionary or preemptive approach rather than a "wait and see" policy to regulation, it is conceivable that many of the innovations that make up modern life today would have been introduced much later, or even not at all: the automobile (with side-effects of more than 40,000 auto related deaths a year), the airplane (which has its share, albeit much lower than the car, of fatalities), and even the Internet (which is used by terrorists and criminals, not just ordinary citizens).

Advocates of preemptive screening in the financial arena no doubt will argue, possibly taking their cue from Paul Volcker's skepticism toward the social value of financial innovation in general, that because the benefits of innovation in the financial arena are likely to be less than in the real sector and the dangers of harmful innovation much greater, financial innovation therefore should be treated differently for regulatory purposes than real sector innovation. My foregoing summary of financial innovations provides a more optimistic assessment than this view. Moreover, the case for preemption wrongly assumes, in my view, that "wait-and-see" regulation cannot be improved.

I believe it can, in part precisely because regulators made mistakes, which they have admitted, in the run-up to the most recent crisis. One of the silver linings of the crisis is that it has demonstrated to elected officials the dangers of interfering with regulatory attempts to clamp down earlier on products and practices that permit asset bubbles to unduly expand and later pop, with devastating consequences. At least for a good long while, regulators—specifically the new Systemic Risk Council of regulators in the United States created by the Dodd-Frank bill—will have not only greater freedom to act, but the *legal duty to do so* to prevent future bubbles, especially those fueled by leverage, from getting out of hand.

Another lesson from the recent crisis is that potentially dangerous bubbles may be forming when particular asset classes or financial instruments are *rapidly growing.* Future regulators would be greatly aided in their efforts to identify and prevent future bubbles and thus possible sources of systemic crises by harnessing important *market-based* signals of distress, such as those provided by the markets in credit default swaps, which regulators can use to justify their own early, preventative actions. Indeed, this potential use of CDSs is an important reason why attempts by regulators in some countries to ban or limit "naked" CDSs or short selling are severely mistaken. They punish the messenger, when in fact we need more market-based messaging to help guide regulators and policy markets.

Finally, notwithstanding the strong general case for a wait-and-see approach to financial regulation, there may be certain aspects of finance where the more intrusive preemptive approach toward innovation may be warranted. One such appropriate area involves financial products involving long-term contracts entered into by consumers, such as mortgages (when borrowing) or annuities (for retirement). There is a growing literature in behavioral finance indicating that individuals are not always rational in their investment decisions. This tendency is dangerous when even well-informed individuals are making long-term financial commitments, with heavy penalties (in the case of mortgages) or perhaps no exit strategies (in the case of annuities) for changing one's mind later. In these cases, preemptive approval of the design of the financial products themselves may be necessary to prevent many consumers from locking themselves into expensive and/or potentially dangerous financial commitments. But this exception should remain that way and not become the rule.

In sum, a balanced look at financial innovation over recent decades reveals a more positive picture than has been painted by some skeptical observers. But regardless of how one assesses past financial innovation, the recent crisis teaches us that policy makers must stand readier to correct abuses when they appear and not let destructive financial innovations wreak the kind of economic havoc that we have unfortunately just witnessed.

BBVA

The Financial Industry and the Crisis: The Role of Innovation

Xavier Vives[*]
IESE Business School

1. INTRODUCTION

The acceleration of the process of liberalization and globalization in the financial sector which began in the United States in the 1970s, initiated and spurred on by changes in information technologies, has not been accompanied by a parallel development of the system's regulatory framework, whose instability has steadily increased. Financial innovation in derivatives and securitization, fuelled by a lax monetary policy, created a bubble in the housing and credit-supply markets which burst when the subprime mortgage crisis hit in 2007. In the past, major technological changes such as the railway, the automobile or the internet have been accompanied by speculative bubbles in a context of asymmetric information and biased predictions, and the effects of financial innovation on derivatives and securitization are no exception to this historical trend.

What were the mechanisms that produced this outcome? How can potential crises be averted or mitigated in the future? Should we impose restrictions on innovation? What role should regulation play?

In order to answer these questions, we must first understand the role of financial innovation in the transformation of banking and the financial markets, determine whether it has increased the fragility and risks of the system, put the contribution of regulation in context, and consider the relationship between economic growth and innovation in the financial industry.

This chapter discusses the role of financial innovation in the transformation of the banking industry (section 2) and in the progress of the crisis (section 3), the effects of asset securitization (section 4) and regulatory reform and the role of agent incentives (section 5), and concludes in section 6.

2. FINANCIAL INNOVATION AND THE TRANSFORMATION OF BANKING

The recent history of the financial sector can be divided into two periods. The first, characterized by strict regulation, interventionism and stability, encompasses the years from the 1940s to the 1970s, while the second was an era of liberalization and growing instability which lasted from the 1970s until 2007, when the subprime mortgage crisis began. The stability of the first period contrasts sharply with the

* Public-Private Sector Research Centre at the IESE Business School. This text was written for inclusion in the book *Innovación. Perspectives for 21st Century* published by BBVA. I am grateful for the excellent research assistance provided by Karla Perca and for the sponsorship of the Abertis Chair of Regulation, Competition and Public Policy.

considerable increase in the number of bankruptcies and crises registered during the second period, when the sector was liberalized. The heightened instability of the latter period has its origin in this liberalization accompanied by a woefully inadequate regulatory framework, as evidenced by the crises in the United States, Japan and Scandinavia.[1] However, and despite these periodic crises, financial liberalization has contributed to the general development of the financial industry and consequently to the growth of the economy.

The liberalization of the financial sector cannot be explained without taking financial innovation into account. To this we must add the progressive globalization of the financial sector and the "shareholder value" movement, which has affected the market for corporate control of banks and companies and has put pressure on banks to obtain higher profitability.

The second period witnessed the advent of numerous innovations in forms of payment (credit and debit cards), transaction processing (ATMs, telephone and online banking, e-commerce for financial assets), saving options (such as investment funds and structured products), loans (automated credit scoring) and risk management techniques (derivatives and securitization). Breakthroughs in information technologies are largely responsible for these new developments which boost productivity, permit a better diversification of risk, and generate economies of scale in internal activities as well as a need for highly qualified and specialized human resources.

Prior to the 2007 crisis, banking had evolved from the traditional business of accepting deposits and granting and supervising loans, to providing services to investors (asset/investment fund management, advice and insurance) and companies (consultancy services, insurance, mergers and acquisitions, underwriting share offerings and debt securities, securitization, risk management), while also engaging in proprietary trading. In a financial conglomerate we can find a retail bank, an investment or merchant bank, asset management, proprietary trading, and insurance. The now-infamous "originate-and-distribute" banking model is a good example of the banking industry's process of evolution. At the same time, although banks created off-balance-sheet entities (SIVs, ABCP conduits), these were guaranteed by liquidity lines.

New developments in information technology have intertwined intermediaries and the financial markets almost inextricably. The importance of a bank's investment portfolio at market value has increased substantially because there are now more opportunities for trading assets, which means that the risk profile of a financial institution can change in a matter of seconds with financial market transactions (for example, using e-commerce and derivatives). The banking industry has increased its market funding, particularly in short-term funds that can be liquidated very quickly. As a result, banking is now more vulnerable to the vicissitudes and volatility of the market, herd-behaviour phenomena, and asset price boom-bust cycles.[2] This in turn increases the risk of illiquidity. Meanwhile, agents may have even greater incentives to take excessive risks that remain hidden from investors—risks that are significant but quite unlikely to materialise (tail risk) due to compensation schemes based on the short-term results achieved by other agents.[3] The effective compensation received by agents, with the approval of the financial intermediaries' shareholders,

[1] See Reinhart and Rogoff (2009) and the analysis in section 2 of Vives (2010a).

[2] See chapters 6 and 8 in Vives (2008).

[3] In addition, if investors demand financial instruments with guaranteed returns and are unaware of the improbable or tail risk, there will be an excess of share offerings and the market will become fragile when the investors finally understand the risks involved (see Gennaoli *et al.*, 2010).

tends to soar when things are going well, and is more inflexible when they are not (in technical terms, it is markedly convex), thus providing an incentive to take excessive risks. Paradoxically, an increase in market depth may be accompanied by a significant rise in systemic risk (Rajan, 2006). The progress of the present crisis is a perfect example.

3. THE COURSE OF THE CRISIS AND REGULATION

In the current financial crisis, the contagion spread and was exacerbated via market channels. The globalization of the financial markets can lead to greater diversification, but it also increases the likelihood of domino-effect contagion between entities and contagion due to information difficulties. The result was the collapse of the asset-backed commercial-paper market (via securitizations) and of the interbank market. Wholesale funding made the situation even more fragile and revealed itself as a critical weakness of the balance sheets of financial institutions, two cases in point being Northern Rock and Lehman Brothers (Shin, 2009; Adrian and Shin, 2010). Leverage evolved procylically with fair value accounting. When asset values rose, the balance sheets of the institutions were strengthened, which in turn allowed them to increase their debt levels, and new asset purchases fuelled the upward climb of prices and leverage. This process was inverted during the second stage of the crisis when de-leveraging began.[4]

> "Prior to the 2007 crisis, banking had evolved from the traditional business of accepting deposits and granting and supervising loans, to providing services to investors (asset/investment fund management, advice and insurance) and companies (consultancy services, insurance, mergers and acquisitions, underwriting share offerings and debt securities, securitization, risk management), while also engaging in proprietary trading"

However, at the epicentre of this crisis was the originate-and-distribute model, which gave rise to an inverted pyramid of complex derivatives based on subprime mortgages. In the originate-and-distribute model, banks try to get rid of credit risk by originating mortgage loans and quickly securitizing them with a chain of increasingly complex structured products. The problem with this model is that it leaves mortgage monitoring in limbo, it is opaque and, given the complexity of the products, it leads to an underestimation of true risk levels. Moreover, the mortgage risk reappears on the bank's balance sheet when its structured investment vehicles (SIVs) begin to experience liquidity problems owing to the institution's explicit and implicit obligations. Risk underestimation was further aggravated by the use of

[4] See chapter 2 in EEAG (2009).

"At the epicentre of this crisis was the originate-and-distribute model, which gave rise to an inverted pyramid of complex derivatives based on subprime mortgages"

statistical models based on short series and historical correlations (and fat-tailed distributions) without taking into account the systemic risk generated by the new products and high levels of leverage. Mechanical risk assessment models that only work within very strict parameters were routinely misused. The opacity of the new derivatives (partially attributable to over-the-counter or (OTC) transactions, which make it difficult to provide a comprehensive assessment of counterparty risk) led to an underestimation of the tremendous systemic risk that had built up in the market as well as to a very serious problem of adverse selection, given that no one knew when the crisis would hit or what the magnitude or distribution of exposure to toxic products derived from subprime mortgages would be. This problem of asymmetric information paralysed the interbank markets, making them illiquid.

A chain of misaligned incentives culminated in catastrophe. Public agencies in the United States encouraged the granting of subprime mortgages to families with limited ability to repay the loans; the credit scoring agencies, siding with securities issuers, vied to see who could give the most favourable scores to the riskiest products; and the short-term compensations available to financial agents led many to take excessive risks (this is true of the originators and distributors of complex products as well as of the buyers). This chain thrived on the incredibly low interest rates that financed the real estate bubble. Meanwhile, monetary policy only concerned itself with inflation, ignoring the bubbles in asset prices and the balance-sheet situation of financial institutions.

The crisis was brought about and exacerbated by inadequate regulation. The first major flaw in regulation was a dualist framework that permitted regulatory arbitrage between the regulated sector of depository institutions and the parallel banking system of structured vehicles and investment banking. The second shortcoming was qualitatively and quantitatively insufficient capital requirements. These low levels of capital were compounded by low liquidity, rendering the system more fragile, while leverage continued to rise. To make matters worse, because capital ratios remained fixed they accentuated the cycle instead of modulating it. In addition, fair value accounting evidenced procyclical tendencies in the leverage cycle. Financial regulations failed to take systemic risk into account, regulators were not properly informed of that risk, and potentially-systemic institutions were not given special treatment. The opacity of the parallel banking system and the unorganized OTC-derivatives markets helped to camouflage the underlying systemic risk. Finally, the important role played by credit scoring agencies in the field of regulation (for example, in determining capital requirements) was reduced to a competition to see who could lower their standards faster, without the proper supervision of any regulatory authority.

> “Spreading the credit risk among investors with different risk profiles facilitates a more efficient use of capital, and banks acquire an additional source of funds which allows them to extend more credit”

4. THE EFFECTS OF ASSET SECURITIZATION

In light of the pivotal role that asset securitization has played in the current crisis, it would seem that an analysis of its benefits and disadvantages is in order. There is little doubt that securitization has facilitated the development of financial markets, permitted credit expansion and contributed to economic growth. However, the recent financial crisis has exposed the weaknesses of this innovation, such as the incentives to over-expand credit by compromising on loan quality, or the complexity of the structured products derived from those loans, which made it hard for investors to evaluate the risks to which they were exposed. The result was a substantial yet hidden increase in systemic risk.

Credit Expansion and Regulatory Arbitrage

By means of securitization, banks can turn illiquid loans, such as mortgages, into tradable instruments. Spreading the credit risk among investors with different risk profiles facilitates a more efficient use of capital, and banks acquire an additional source of funds which allows them to extend more credit. At the same time, securitization makes it possible to reduce their legally-stipulated capital requirements by selling the loans to off-balance-sheet vehicles. These loans can be entirely dissociated from the originating institution or not in order to lower capital requirements.[5] Naturally, the ability to maintain a high level of credit supply with less capital allowed banks to cut financing costs for loan recipients and offered people who would not normally be considered creditworthy the chance to take out mortgages (and other types of loans).[6]

Loan Quality Deterioration

The originate-and-distribute model gave rise to the application of laxer criteria when selecting loan recipients and fewer incentives to monitor borrowers. The ability to quickly shift at least part of the risk onto other investors by using structured products, coupled with the assumption that mortgage refinancing was always possible given the steady rise of housing prices, resulted in the application of lower standards for evaluating the default risk of loan recipients. This situation increased the level of risk in the entire financial system (Keys *et al.*, 2008).

Higher Systemic Risk

Securitization allows banks to redistribute risk to those investors most willing to bear it. However, when evaluating the diversification potential of securitization risk, one must bear in mind that lower diversifiable risks increase the level of systemic risk. Thus, when faced with an event which negatively affects the economy as a whole, such as plummeting housing prices, structured products will be harder hit than traditional instruments with the same credit rating (Colval *et al.*, 2008). Meanwhile, liquidity risk also rose and

[5] Under the terms of Basel I, by selling loans to off-balance-sheet vehicles, banks were able to reduce the capital they needed to meet regulatory requirements. In the Basel II framework, banks could transfer loans to off-balance-sheet vehicles and endow them with liquidity lines, turning them into instruments with the highest possible rating (triple-A). In this way, banks could then buy back these instruments and include them in their balance sheets, thus lowering their capital requirements. (See Brunnermeier (2009) and chapter 2 in EEAG (2009))

[6] See ECB (2008). Sabry and Okongwu (2009), for example, show that securitization in the United States has resulted in an increased availability of credit and lower loan costs. Between 1999 and 2006, a 10% increase in the level of securitization led to a decrease of between 4 and 64 basis points in loan yield spreads, depending on the type of loan analysed (mortgages, car loans and credit card loans).

contributed significantly to systemic risk, because off-balance-sheet vehicles were funded by commercial-paper issuance which was backed by long-term mortgages but had short- or medium-term maturities (average of 90 days and one year, respectively). Thus, the principal and the interests were paid in part with the cash flow generated by mortgages, and the rest was paid by issuing new securities. Banks further increased this risk by providing their vehicles with liquidity backstops to safeguard against any temporary inability to pay investors.[7]

Finally, structured products derived from loans, which were often granted without considering the credit risk, are hard to evaluate. The structure of these products—built upon a portfolio of loans which is subsequently divided into tranches with different risk/return profiles, and is usually restructured into new complex securities (re-securitization via collateralised debt obligations or CDOs)—can ultimately result in a lack of information about the risks to which investors are exposed, given their distance from the underlying loans, and making direct assessment virtually impossible. This opacity derived from the securitization process is considered a crucial factor in the loss of confidence in the financial system, which ended up triggering the crisis.[8]

Credit Rating Agencies and Complexity

Given the complexity of structured products, investor purchase decisions were largely based on the ratings provided by risk assessment agencies. The subprime mortgage crisis revealed two major problems in this area. Firstly, the same rating scale was applied to structured and traditional products, yet one of the things that characterizes structured products is their ability to transform risky loans into highly-

> "Given the complexity of structured products, investor purchase decisions were largely based on the ratings provided by risk assessment agencies"

rated instruments by creating tranches according to priority of payment, targeting investors with different risk profiles.[9] In this way, investors could purchase products with the best possible rating but which offered a higher yield than traditional bonds. Moreover, the banks made sure that payment tranches were designed in such a way that they just barely met the minimum requirements for AAA rating (a practice known as *rating at the edge*). Secondly, investors did not account for the fact that credit ratings were based on calculations which only considered default risk and ignored the risk that the ratings themselves could be revised downwards or that the situation of the housing market could change (IMF, 2008). Another factor that contributed to the favourable rating of structured products in comparison to traditional bonds is the fact that rating agencies charged the issuers higher commissions for structured products.

5. REFORMING REGULATION AND INCENTIVES

Like any technological breakthrough, financial innovation can either improve the economy's efficiency or introduce activities that generate private benefits as well as social costs (negative externalities). Innovations that enhance markets,

[7] See chapter 2 in EEAG (2009) and Brunnermeier (2009).

[8] See, for example, Gorton (2008), Brunnermeier (2009) and chapter 2 in EEAG (2009).

[9] Approximately 75% of all subprime mortgages in the United States have been securitized. Of this percentage, 80% were funded by tranches of senior or AAA-rated securities (IMF, 2008).

providing financial instruments that offer new possibilities of diversification and risk coverage (such as options and futures), and help them overcome problems of asymmetric information (the typical debt contract, for example) are beneficial. Examples of the second possibility include financial instruments that facilitate rent seeking, taking advantage of investors or consumers through obfuscation, the inflation of speculative bubbles, the increasing fragility of the system, and regulatory arbitrage when adequate regulation exists. Following the advent of the crisis, prominent economists and public decision-makers (Paul Volcker, Lord Turner, Paul Krugman, Simon Johnson and James Kwak, to name but a few) voiced their scepticism about the positive contributions of financial innovation. Nevertheless, it is obvious that many financial innovations have boosted economic growth, and the relationship between financial progress and economic progress is well documented (Levine, 2005). We should also remember that financial innovation (venture capital, for example) has played an important part in the development of new technologies and innovative firms in a variety of sectors.[10]

How innovation is used is determined by the incentives of the economic agents, who are in turn influenced by the regulatory framework. For example, some analysts are now exploring the degree to which pressure to generate value for shareholders and possible flaws in corporate governance mechanisms contributed to the crisis. The limited responsibility of shareholders in a context of deposit guarantees and explicit or implicit TBTF (too big to fail) policies leads investors to demand high-risk options, given that the profits are private and the losses, in the event of bankruptcy, are shouldered by society to a large extent. Shareholders

“Like any technological breakthrough, financial innovation can either improve the economy’s efficiency or introduce activities that generate private benefits as well as social costs (negative externalities)”

therefore agree to compensation contracts for executives that encourage risk-taking, with a remuneration package that is unaffected when share prices drop but shoots up when they rise. There is recent evidence which indicates that this occurred in the pre-crisis period.[11] Of course, there can also be additional problems of agency (conflicts of interest) between shareholders and executives and between executives and the financial intermediaries’ traders.

Therefore, the main issue is actually incentives and reforming the regulatory framework so that private agents shoulder the potential social costs of their decisions. The regulatory reform now underway will be successful if it embraces the following principles: the existence of a systemic risk regulator; standardized regulations for all entities that provide banking services (to avoid regulatory arbitrage); risk premiums and limited scopes of activity in keeping with the characteristics of each intermediary; capital requirements and rates that take systemic risk into account; and a holistic approach that brings the incentives of the

[10] See Litan (2009) for a defence of many financial innovations.

[11] See Fahlenbrach and Stulz (2009), Cheng *et al.* (2010), Bebchuk and Spamann (2010), and Bebchuk, Cohen and Spamann (2010). The words of Chuck Prince, CEO of Citigroup (*Financial Times,* July 2007), can also be interpreted in this light: “When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you’ve got to get up and dance. We’re still dancing.”

> “Derivatives markets provide economic agents with opportunities for risk coverage and signposts that condense the scattered information floating around the market, and this role can be maintained with trading in organized markets, monitoring, and transparent information on counterparty risk”

system's various agents into line, both domestically and internationally.

The process of reforming liquidity and capital requirements (known as Basel III) and the legislative reforms introduced in the EU and the USA are headed in the right direction, though they may have limitations; however, since these reforms have not yet taken root, it is still too soon to determine whether or not they will be sufficient (Vives, 2010b). For example, the Dodd-Frank Act passed in the United States in July 2010 has established a variety of measures to align private and social incentives in innovative products or markets. Banks wishing to complete derivative transactions must now go through central clearing instead of engaging in direct OTC transactions, which are under federal supervision. Among other things, this regulation seeks to prevent a cascade of losses in the event of failure of a major player in the OTC market of credit default swaps (CDS), which offers protection against potential default on a loan or bond. The act also establishes prudential standards and rules on transparency, designed to help the securitization market recover its pivotal role in funding the economy. For example, originators are now required to retain part of the credit risk (5%), giving them a good incentive to monitor loans. In addition, the law created a consumer protection agency to help restore investor confidence and overcome the conflicts of interest that have infested the financial industry. This agency may be instrumental in improving transparency for consumers and investors, facilitating the comparison of financial products and services offered by different companies, and curtailing the deleterious effects of innovations that increase opacity.

However, there are some questionable aspects of the regulatory reform. The proposed reforms for corporate governance in the financial sector run the risk of being ineffective if they fail to address the root problem of the incentives generated by deposit insurance and the bailouts of TBTF institutions which, combined with limited responsibility, induce shareholders to take risks which are excessive from a social standpoint. With regard to market reform, the desirability of the restrictions on short selling or naked shorting imposed in certain countries is questionable given that the root problem is market manipulation.

6. CONCLUSION

Financial innovation has been accused of destabilising the banking industry and the financial markets and of helping operators get around regulatory requirements. Although these accusations are true in some cases (such as the abuse of certain complex structured products), the real underlying problem is not innovation per se but inadequate regulation. For example, derivatives markets provide economic agents with opportunities for risk coverage and signposts that condense the scattered information floating around the market, and this role can be maintained with trading in organized markets, monitoring, and transparent information on counterparty risk. Securitization is an innovation that allows investors to transfer risk and diversify, which in turn increases the amount of available credit in an economy. The problems that have been detected derive from a chain of inappropriate incentives in a context of deficient regulation.

Innovation is necessary for the progress of the financial system, and this progress is an essential ingredient for economic growth. The challenge is to devise a regulatory framework which allows innovation, globalization and the financial system to develop while ensuring a proper balance between private and social incentives.

BIBLIOGRAPHY

Adrian, T., and H. S. Shin (2010), "Liquidity and Leverage", *Journal of Financial Intermediation* 19(3), pp, 418-437.

Bebchuk, L., A. Cohen and H. Spamann (2010), "The Wages of Failure: Executive Compensation at Bear Stearns and Lehman 2000-2008", *Yale Journal on Regulation* 27(2), pp. 257-282.

Bebchuk, L., and H. Spamann (2010), "Regulating Bankers' Pay", *Georgetown Law Journal* 98(2), pp. 247-287.

Brunnermeier, M. (2009), "Deciphering the Liquidity and Credit Crunch 2007-2008", *Journal of Economic Perspectives* 23(1), pp. 77-100.

Cheng, I.-H., H. Hong and J. Scheinkman (2010), "Yesterday's Heroes: Compensation and Creative Risk-Taking", mimeo.

Colval, J., J. Jurek and E. Stafford (2008),"The Economics of Structured Finance", Working Paper 09-060, Harvard Business School.

EEAG (2009), "The EEAG Report on the European Economy 2009", Munich: CESifo.

European Central Bank (2008), "Monthly Bulletin", February 2008.

Fahlenbrach, R., and R. Stulz (2009), "Bank CEO Incentives and the Credit Crisis**",** ECGI Finance Working Paper 256.

Gennaioli, N., A. Shleifer and R. Vishny (2010), "Financial Innovation and Financial Fragility", NBER Working Papers 16068.

Gorton, G. (2008), "The Subprime Panic", Yale ICF Working Paper no. 08-25.

International Monetary Fund (2008), "Global Financial Stability Report: Containing Systemic Risks and Restoring Financial Soundness", April.

Keys, B., T. Mukherjee, A. Seru and V. Vig (2008), "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans", EFA 2008 Athens Meeting Paper.

Levine, R. (2005), "Finance and Growth: Theory and Evidence", in P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, Amsterdam: Elsevier/ North-Holland.

Litan, R. (2010), "In Defense of Much, But Not All, Financial Innovation",The Brookings Institution Working Paper.

Rajan, R. (2006), "Has Finance Made the World Riskier?", *European Financial Management* 12(4), pp. 499-533.

Reinhart, C., and K. Rogoff (2009), *This Time is Different: Eight Centuries of Financial Folly*, Princeton: Princeton University Press, p. 496.

Sabry, F., and C. Okongwu (2009), "Study of the Impact of Securitization on Consumers, Investors, Financial Institutions and the Capital Markets", NERA Economic Consulting.

Shin, H. S. (2009), "Reflections on Northern Rock: The Bank Run That Heralded the Global Financial Crisis", *Journal of Economic Perspectives* 23(1), pp. 101-119.

Vives, X. (2008), *Information and Learning in Markets*, Princeton: Princeton University Press.

Vives, X. (2010a), "Competition and Stability in Banking", CESifo WP 3050, soon to be published in *Monetary Policy under Financial Turbulence, Proceedings of the Annual Conference of the Central Bank of Chile.*

Vives, X. (2010b), "La crisis financiera y la regulación", soon to be published in *Mediterráneo Económico.*

BBVA

Innovation and Climate Change

Edward S. Rubin
Carnegie Mellon University

INTRODUCTION

To a large extent, the study of innovation and technological change has been motivated by a desire to understand and shape the forces that underlie economic development and competitiveness in a market economy. Thus, there is a large literature, contributed mainly by social scientists, examining the many facets of innovation and the factors that contribute to it—ranging from the behavior of individuals and organizations, to the role and effectiveness of government policies aimed at spurring innovation in particular sectors of the economy or targeted areas of technology such as computers, aircraft, or agriculture.

The role of technological innovation in addressing societal problems such as air pollution and water pollution is a more recent development. Unlike innovations in industries such as pharmaceuticals or electronics—where the result is new products that consumers desire (such as more effective or lower-cost medicines, cell phones and internet services)—there is little or no "natural" market for most environmental technologies whose function is to reduce or eliminate a pollutant discharge to the environment. Would you voluntarily pay an extra $1,000 to install air pollution emission controls on your automobile if it were up to each consumer to decide? Most individuals would not, recognizing that their action alone would do little to solve the air pollution problem unless all drivers were required to take the same action.

In cases such as this, the role of government policies and regulations becomes critical, since most environmental problems require collective action to effectively address the problems. Similarly, the nature and extent of innovations that lower the cost and/or improve the efficiency of environmental controls depends heavily on the actions of government agencies at all levels.

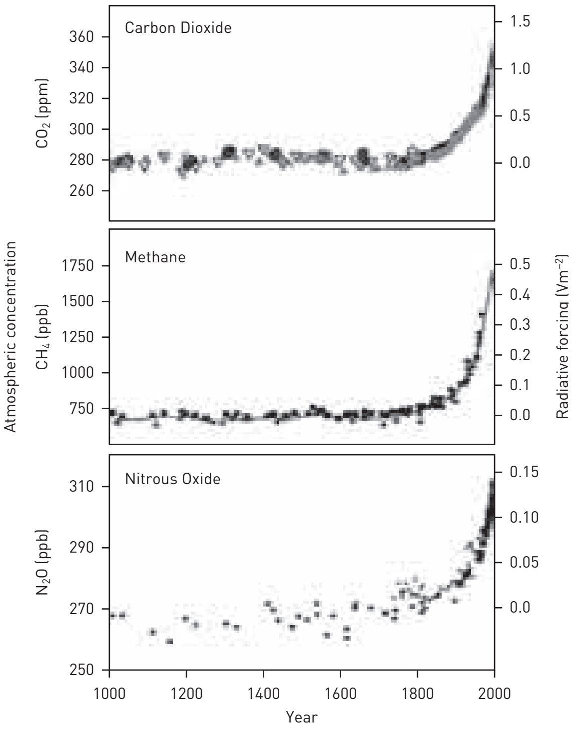
In this paper we focus on the links between technological innovation and global climate change—which is arguably the most far-reaching and formidable environmental challenge facing the world today. First we present a brief overview of the climate change problem and the innovation needs that motivate this paper. Then we examine in greater detail some of the options available to accelerate the innovations needed to address the climate change challenge. While many of the examples cited in this paper are drawn from experience and studies for the United States, the general concepts and approaches that are discussed are widely applicable to all nations faced with the challenges of climate change mitigation.

THE CLIMATE CHANGE PROBLEM

Over the past 150 years, there have been significant increases in the concentration of "greenhouse gases" (GHGs) in the atmosphere, notably carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (see figure 1), as well as a group of industrial GHGs including hydrofluorcarbons (HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride (SF_6). Greenhouse gases drive climate change by trapping heat in the atmosphere, which tends to raise the average temperature of the planet. This, in turn, alters the patterns and intensity of precipitation as well as the flows of air and ocean currents around the globe—all of which directly or indirectly influence the climate (defined as the average weather in a region over a period of several decades.)

The main sources of increased GHGs in the atmosphere are the GHG emissions from a variety of human activities (table 1).

Figure 1. Historical trend in the atmospheric concentration of major GHGs



Source: IPCC, 2007a

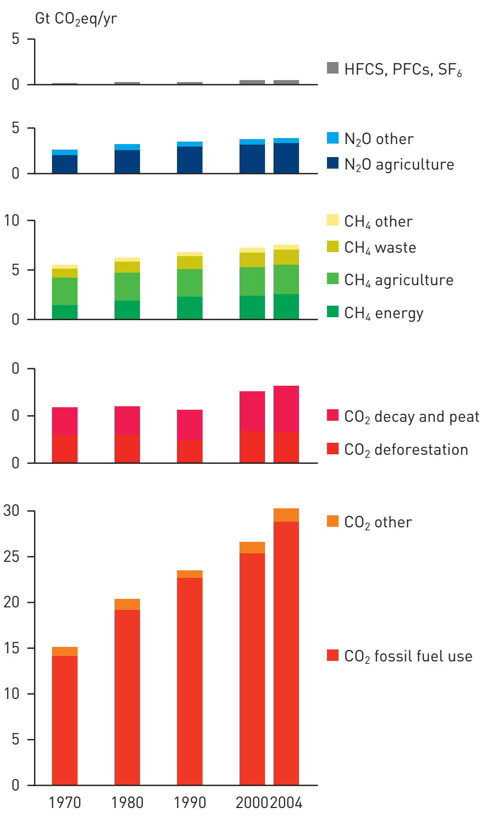
Figure 2 shows the recent growth in global GHG emissions, expressed in terms of "CO_2 equivalent" tonnages, which accounts for differences in the heat-trapping ability of different gases relative to carbon dioxide (see IPCC, 2007 for details). The largest contributor is CO_2 from the combustion of fossil fuels (petroleum, coal, and natural gas,

Table 1. The major greenhouse gases and common sources of emissions

Símbolo	Nombre	Fuentes comunes
CO_2	Carbon Dioxide	Fossil fuel combustion, forest clearing, cement production, etc.
CH_4	Methane	Landfills, production and distribution of natural gas & petroleum, fermetation from the digestive system of livestock, rice cultivation, fossil fuel combustion, etc.
N_2O	Nitrous Oxide	Fossil fuel combustion, fertilizers, nylon production, manure, etc.
HFC's	Hydrofluorocarbons	Refrigeration gases, aluminium smelting, semiconductor manufacturing, etc.
PFC's	Perfluorocarbons	Aluminium production, semiconductor industry, etc.
SF_6	Sulfur Hexafluoride	Electrical transmissions and distribution systems, circuit breakers, magnesium production, etc.

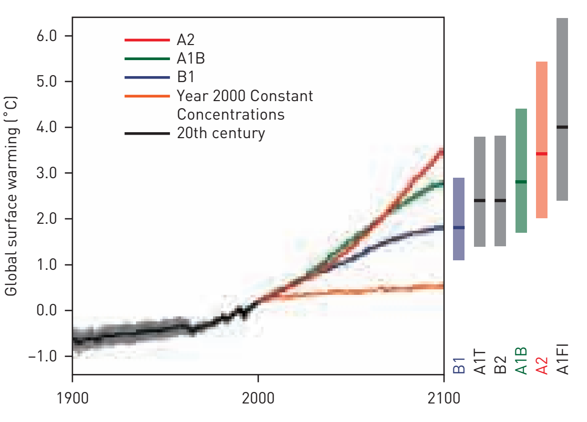
Source: IPCC, 2007b

Figure 2. Recent growth in global emissions of greenhouse gases



Source: IPCC, 2007b

Figure 3. Historical trend and future scenarios of global warming from 1900 to 2100. Ranges shown at the right are for six scenarios (labeled B1 through A1FI) modeled by the IPCC



Source: IPCC, 2007b

composed mainly of carbon and hydrogen). Because our use of energy also releases some non-CO_2 GHGs (primarily CH_4 and N_2O), energy use accounts for roughly 85 percent of all GHG emissions.

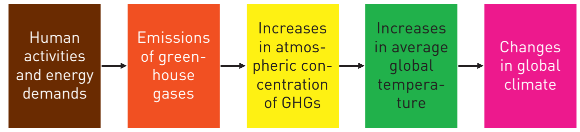
The essence of the climate change problem is that if current trends continue, future global emissions of greenhouse gases will grow significantly in coming decades in response to growth in world population, economic development, and other factors that increase GHG emissions. As a result, the average global temperature is projected to increase by 1.1°C to 6.4°C by the end of this century (IPCC, 2007). While there is considerable uncertainty in such projections (as evidenced by figure 3), the potential impacts of global warming could seriously endanger human health, water supplies, agriculture, and human settlements—especially in coastal areas vulnerable to sea level rise and storms (IPCC, 2007b; NRC, 2010b).

In light of these large uncertainties, why not simply wait until there is stronger empirical evidence about the magnitude and impacts of climate change? A fundamental difference between greenhouse gases and "conventional" air pollutants like sulfur dioxide (SO_2) and particulate matter is that GHGs, once emitted, remain in the atmosphere for very long periods of time—typically decades to millennia. For example, roughly half the CO_2 emitted today will still be in the atmosphere a century from now, still contributing to global warming. Centuries later some of today's CO_2 emissions will still be in the air! In contrast, conventional pollutants like SO_2 stay in the atmosphere for relatively short periods of time—typically days or weeks—before they are removed or washed out by various physical and chemical processes. Thus, if we quickly reduced emissions of conventional pollutants their atmospheric concentrations (and associated impacts) also would fall quickly. Not so for GHGs. Because of their long lifetimes, atmospheric concentrations would continue to rise unless emissions were curtailed dramatically. (Think of a bathtub being filled from a large faucet, with only a slow trickle draining from the bottom; the water level would continue to rise unless the faucet were turned down nearly all the way to match the slow drainage.) Thus, if climate change impacts prove to be as serious as projected, reducing GHG emissions in the future would do little to quickly reduce atmospheric concentrations to mitigate those harmful impacts.

WHAT ACTIONS ARE NEEDED?

International policy goals for global climate change were established in 1992 under the United Nations Framework Convention on Climate Change (UNFCCC). To date, 192 nations have adopted the UNFCCC goal of "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system". Scientific research has sought to better understand and quantify the links between human activities, GHG emissions, the resulting increases in atmospheric concentration, the consequent changes in global temperature, and the impacts of those changes (figure 4). The largest uncertainties are in the links between global temperature increases and resulting impacts. However, based on current science many policymakers worldwide advocate no more than a 2°C rise in long-term global temperature as the climate policy goal needed to prevent dangerous impacts. Achieving that goal would require actions to stabilize atmospheric GHG

Table 2. Key linkages between human activities and global climate change



concentrations at levels only slightly greater than current levels. That, in turn, would require a reduction in annual global GHG emissions of 50% to 80% below 1990 levels by 2050, according to recent studies (IPCC, 2007b).

The technological implications and challenges of meeting such a goal are formidable. This is illustrated in figure 5, which shows the results of recent modeling studies for the United States. These results show there is no unique solution or pathway to achieving large reductions in GHG emissions—different models give different solutions based on different assumptions about the future availability and cost of alternative technologies and other factors. What all models show emphatically, however, is that dramatic changes in the energy system will be required, since this is the dominant contributor to climate change.

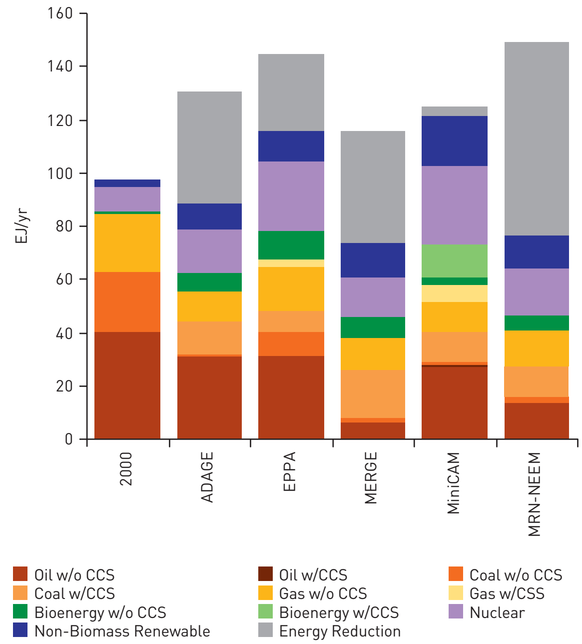
Today about 85% of the world's energy is provided by fossil fuels. Approximately half of that is in the form of oil (used mainly for transportation), followed by roughly equal amounts of coal (used primarily for electricity generation) and natural gas (used for a variety of domestic and industrial heating applications, and increasingly for electric power generation). The CO_2 released from the combustion of those fuels—primarily from power plants and automobiles—is the key source of GHG emissions. Achieving a transition to a sustainable low-carbon (ideally zero-carbon) energy system is the major challenge we face to avoid potentially dangerous climate change.

THE NEED FOR TECHNOLOGICAL CHANGE

Technological change on a massive scale will be needed to achieve large reductions in global GHG emissions. The results in figure 5 illustrate the four general strategies available to transform the energy system of a country or region: 1. reduce the demands for energy in all major sectors of the economy (buildings, transportation, and industry), thus reducing the demand for fossil fuels; 2. improve the efficiency of energy utilization so that less fossil fuel is required to meet "end use" energy demands, resulting in lower CO_2 emissions; 3. replace high-carbon fossil fuels such as coal and oil with lower-carbon or zero-carbon alternatives such as natural gas, nuclear, and renewable energy sources such as biomass, wind and solar; and, 4. capture and sequester the CO_2 emitted by the combustion of fossil fuels to prevent its release to the atmosphere.

As illustrated for the scenario in figure 5 (an 80% reduction below 1990 emissions by 2050), all four approaches are needed to reduce emissions at lowest cost. Reductions in energy demand, which include the effects of improved efficiency, play the most prominent role in all but one of the five models shown. The uncontrolled combustion of coal is eliminated or sharply curtailed in all cases, and the direct use of oil and natural gas also is reduced relative to the year 2000 reference case. In contrast, the use of nuclear power, biomass, and non-biomass renewables (mainly wind) increases significantly in these studies. So too does the use of carbon capture and storage (CCS). This technology could make it possible to capture the CO_2 from power plants and other large industrial sources, and then sequester

Figure 4. Results from five models showing the least-cost US energy mix in 2050 for a policy scenario requiring an 80% reduction in GHG emissions below 1990 levels. Actual energy use in 2000 is shown for reference. (Note: CCS = carbon capture and storage).



Source: adapted from Fawcett *et al.*, 2009

it in deep geologic formations or depleted oil and gas reservoirs. This option has gained substantial worldwide attention in recent years, with efforts now underway to develop and demonstrate the applicability of CCS for climate change mitigation.

The same types of energy system transformations that are illustrated in figure 5 for the United States emerge in other modeling studies at the global level (e.g., IPCC, 2007b; Clark *et al.*, 2009). While energy use is the dominant contributor to GHG emissions, technological change in other sectors will also be needed to deal effectively with climate change. For example, changes in land-use practices, especially deforestation, are needed to reduce or prevent the release of CO_2 from natural "sinks" such as forests

> “The development and adoption of new technology is an essential element of any comprehensive response to global climate change”

and soils. Technological change similarly can reduce or avoid emissions of non-CO_2 GHGs, such as PFCs in the semiconductor industry or nitrous oxide emissions from the agricultural sector. More broadly, at least some adaptation to climate change will almost certainly be necessary, and such adaptations also will require some degree of technological change (NRC, 2010c).

In short, the development and adoption of new technology is an essential element of any comprehensive response to global climate change. But technological change on the scale required cannot happen overnight. To achieve the substantial reduction in CO_2 emissions underlying figure 5, for example, the United States alone would have to retrofit or replace hundreds of electric power plants, tens of millions of vehicles, and hundreds of millions of consumer appliances, building systems (for heating, cooling and lighting), and industrial processes and equipment. Change on this scale will take many decades to achieve.

Many of the technologies needed do not yet exist commercially or are too costly (alternatives to gasoline-powered automobiles is a good example). Some alternatives, such as carbon capture and sequestration technologies for power plants, have yet to gain widespread social and political acceptance. Because the rates of development and adoption of new technologies respond to government policies as well as to market forces such as energy prices, we next look more closely at the processes of technological change and innovation and the factors that influence them.

THE PROCESS OF TECHNOLOGICAL CHANGE

As discussed elsewhere (e.g., NRC, 2010a), the general process of technological change can be characterized as involving a number of steps or stages. Different terms are used in the literature to describe these stages, but four commonly used descriptors are:

Invention: Discovery: the creation of new knowledge or new prototypes;

Innovation: Creation of a new or improved commercial product or process;

Adoption: Initial deployment and use of the new technology;

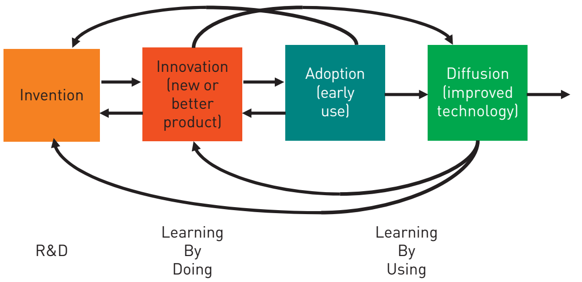
Diffusion: Increasingly widespread adoption and use of the technology.

The first stage—invention—is driven in large part (but not solely) by research and development (R&D), including both basic and applied research. The second stage—innovation—is a term often used colloquially to describe the overall process of technological change. As used here, however, it refers only to the creation of a product or process that is commercially offered; it does not mean the product will be adopted or become widely used. That happens only if the product succeeds in the final two stages—adoption and diffusion, which reflect the commercial success of a technological innovation. Those two stages are the ones that inevitably are most critical to reducing GHG emissions via technological change.

Studies also show that rather than being a simple linear process in which one stage follows another, the four stages of technological change are highly interactive, as depicted in figure 6. Thus, innovation is stimulated not only by R&D, but also by the experience of early adopters, plus added knowledge gained as a technology diffuses more widely into the marketplace. Thus, "learning by doing" (economies in the manufacture of a product) and "learning by using" (economies in the operation of a product) are often (though not always) critical elements that enable the adoption and diffusion of new technologies. Along with sustained R&D (sometimes called "learning by searching"), these stages often help to improve the performance and/or reduce the cost of a new technology—trends that are commonly characterized and modeled as a "learning curve" or "experience curve" (IEA, 2000; McDonald and Schrattenholzer, 2002).

Each stage of the process also requires different types of incentives to promote the overall goal of technological change. An incentive that works well at one stage of the process may be ineffective—or even counterproductive—at another. Large-scale change also must be viewed and considered from a "systems" perspective since the success of any new technology is often dependent upon other technological and non-technological factors. For example, the diffusion of energy-saving technologies that can automatically adjust home appliances like air conditioners and water heaters may depend on the development and dissemination of "smart grid" technology in electrical networks. Similarly, the dissemination of energy-efficient appliances may be inhibited by institutional arrangements, such as landlord-tenant relationships where neither party has

Table 3. Stages of technological change and their interactions



Source: Rubin, 2005

an incentive to purchase a more expensive but more energy-efficient appliance. Thus, in addition to technical considerations, the widespread adoption and dissemination of a new technology may require measures to address social and institutional barriers that affect the nature and pace of technological change.

THE IMPORTANCE OF TECHNOLOGICAL INNOVATION

Any successful strategy to reduce GHG emissions significantly will require actions not only to deploy the low-emission technologies that are available today, but also to foster innovation on new technologies that are needed. Accordingly, there has been growing interest in recent years on ways to foster such innovation, in particular, the role that governments can and should play in that process.

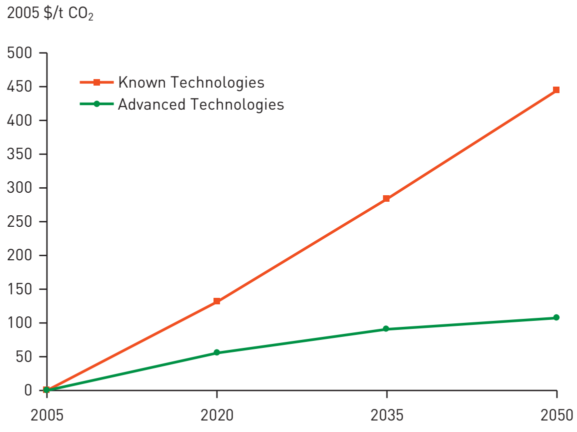
Although research and development is a major element of the innovation process, there is growing recognition that technological innovation is a complex process that commonly involves interactions with other stages of technological change, as depicted in figure 6. Thus, gains from new technologies often are realized only with widespread adoption—a process that usually takes considerable time (often decades) and typically involves a sequence of incremental improvements that enhance performance and reduce costs (Alic *et al.*, 2003).

In the context of this paper, a key question is: what strategies and policies can most effectively foster technological innovations that help reduce GHG emissions? As discussed earlier, GHG emissions depend mainly on the types of energy sources and technologies used to provide the goods and services that society seeks. Thus, technological innovations can help reduce GHG emissions in a variety of ways (NRC, 2010). For example:

- New or improved technologies can enable devices such as vehicles, machinery and appliances to use energy more efficiently, thereby reducing their energy use and GHG emissions per unit of useful product or service (such as a vehicle-mile of travel or a lumen of lighting for illumination).
- New technologies can create or utilize alternative energy carriers and chemicals that emit less GHG per unit of useful product or service (such as renewable energy sources or new low-nitrogen fertilizers).
- New technologies can create alternative ways of providing goods and services that are less GHG-intensive (such as by using substitute products or materials that have lower GHG emissions, or by facilitating larger system-wide changes such as replacing automotive and air travel with teleconferencing and telecommuting).

Technological innovations can facilitate this full spectrum of possibilities. An even broader set of innovations would include social and institutional systems and designs. For example, innovations in urban planning and development could help reduce future energy demands (and associated GHG emissions) for transportation as well as

Figure 5. Global carbon (CO_2-equiv) prices needed to reduce emissions from fossil fuel use and industrial sources with and without advanced technologies



Source: Kyle *et al.*, 2009.

in residential and commercial buildings. Institutional innovations could provide incentives for electric utility companies and others to invest in measures that reduce the demands for energy, as opposed to policies that favor increased energy sales.

Figure 7 shows one estimate of how technological innovations can reduce the future cost of reducing GHG emissions. In this modeling study, a "business as usual" case—which includes historical rates of technological improvements—is compared to a case with more rapid technological change. The cost of meeting a stringent emission-reduction scenario is reduced dramatically when "advanced technologies" are available. This reduction in the unit cost of abatement translates into large national and global cost savings, especially as emission-reduction requirements grow more stringent over time.

THE CRITICAL ROLE OF GOVERNMENT POLICY

A major challenge in reducing GHG emissions is that few if any markets exist for many of the more efficient and low-emission technologies that are needed. What electric utility company, for example, would want to spend a large amount of money on carbon capture and storage technology if there is no requirement or incentive to significantly reduce CO_2 emissions? How many individuals would willingly buy an advanced electric vehicle that costs much more than a conventional automobile simply to reduce their carbon footprint? Costly actions by firms or individuals to reduce their GHG emissions provide little or no tangible value to that firm or person. Only by government actions that either require or make it financially worthwhile to reduce GHG emissions are sizeable markets created for the products and services that enable such reductions. Government actions to create or enhance markets for GHG emission-reducing technologies are thus a critical element of the technological innovation process.

Different policy measures influence technological innovation in different ways. In general, policy options can be grouped into two categories: voluntary measures and mandatory requirements ("carrots" and "sticks"). The first group—often called "technology-policy" options—provides incentives of various types to encourage certain actions or technology developments. The second group consists of government actions that impose requirements or limitations on specified activities, facilities, or technologies, typically in the form of regulations and standards. Table 2 lists examples of policy options in each of these two general categories. The discussions below elaborate briefly on policies in each category to illustrate their role in stimulating innovations that reduce GHG emissions.

TECHNOLOGY-POLICY OPTIONS

Technology-policy measures can stimulate innovation and help create markets for GHG-friendly technologies by providing incentives and support for the development

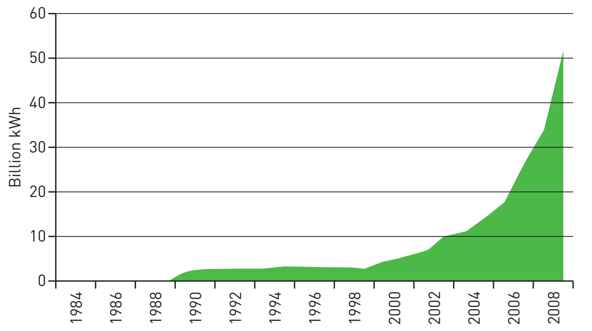
Table 4. Policy options that can foster technology innovations to reduce GHG emissions

"Technology Policy" Options			Regulatory Policy Options
Direct Government Funding of Knowledge Generation	**Direct or Indirect Support for Commercialization and Production**	**Knowledge Dissemination and Learning**	**Economy-wide Measures and Sector or Technology-specific Regulations and Standards**
• R&D contracts with private firms (fully funded or cost shared) • R&D contracts and grants with universities and non-profits • Intramural R&D in government laboratories • R&D contracts with consortia or collaborations	• R&D tax credits • Patents • Production subsidies or tax credits for firms bringing new technologies to market • Tax credits, rebates or payments for purchasers/users of new technologies • Gov. procurement of new or advanced technologies • Demonstration projects • Loan guarantees • Monetary prizes	• Education and training • Codification and dissemination of technical knowledge (e.g., via interpretation and validation of R&D results; screening; support for databases) • Technical standards • Technology/Industry extension programs • Publicity, persuasion and consumer information	• Emissions tax • Cap-and-trade program • Performance standards (for emission rates, efficiency or other measures of performance) • Fuels tax • Portfolio standards

Source: NRC, 2010a

“Studies have documented the ability of energy and environmental regulatory policies to influence the development and deployment of major energy-related technologies, and also to stimulate innovations that reduce GHG emissions and other air pollutants”

Figure 6. Growth in electricity generation from wind in the United States



Source: EIA, 2010

and deployment of new technology. Table 2 lists a number of available measures, grouped into three categories. The first is direct government support for R&D to generate new knowledge (including new concepts and technologies). This is the most common form of government support for innovation, and typically involves a variety of public and private organizations (Alic *et al.*, 2003; CATF, 2009).

The second column lists additional policy options that directly or indirectly support the development, deployment and commercialization of new technologies. Such measures have had a major impact on technology development in the past. For example, US government procurement of jet aircraft and computers during their early stages of commercialization following World War II was critical to their subsequent development and widespread deployment in the marketplace (Alic *et al.*, 2003). More recently, government support in the form of investment tax credits and production tax credits (or feed-in tariffs) have fueled the rapid growth in wind-power systems, as illustrated in figure 8. Additional measures such as loan guarantees and support for demonstration projects are currently being used to stimulate investments in “clean coal” technologies such as coal gasification and carbon capture and storage systems.

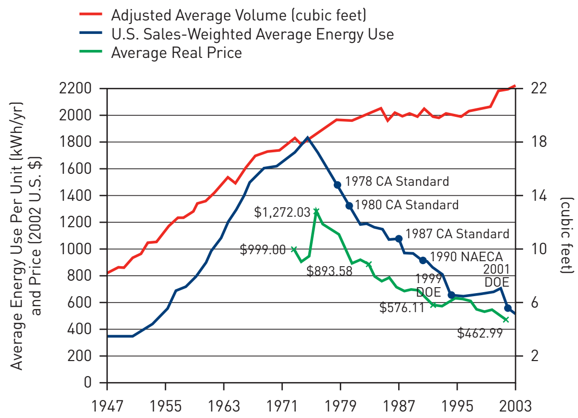
The third group of technology policy options in table 2 reflects measures to stimulate learning and the diffusion of knowledge. These include support for education and training programs, as well as measures such as the development of codes and standards that facilitate the diffusion of new technologies.

REGULATORY POLICY OPTIONS

Energy and environmental regulatory policies respond to “market failures” in which individuals and organizations have little or no economic incentive to curtail activities that adversely affect society as a whole (such as emitting pollutants to the environment), and lack of government intervention. Studies have documented the ability of energy and environmental regulatory policies to influence the development and deployment of major energy-related technologies, and also to stimulate innovations that reduce GHG emissions and other air pollutants. Highly-cited examples include fuel economy and pollutant emission standards for automobiles (Lee *et al.,* 2010), energy efficiency standards for major appliances such as refrigerators (Rosenfeld, 2008), new source performance standards for power-plant air pollutants (Rubin *et al.,* 2004), and market incentives such as the cap-and-trade rules for power plant SO_2 emissions (Popp, 2003).

In 1975, for example, the US government imposed Corporate Average Fuel Economy (CAFE) standards on all new cars sold in the United States in order to reduce US oil consumption in the wake of the 1973 Arab oil embargo. The standards called for roughly a doubling of the average 1973 fuel economy of approximately 13 miles per gallon (mpg) to the CAFE standard of 27.5 mpg for new passenger cars. This provoked a series of technological innovations that affected nearly all aspects of automobile design. In little more than a decade, the US auto fleet became nearly twice as efficient as it had been (EIA, 2010). In 2007, in response to renewed concerns about oil imports, the US adopted more stringent CAFE standards. The new rules call for a fleet-wide average fuel economy (including both passenger cars and trucks) of 34.1 mpg by 2016 (NHTSA, 2010). These standards also will reduce emissions of greenhouse gases (CO_2) from fuel burning. Although the United States has long avoided energy pricing policies and fuel taxes to encourage energy efficiency, evidence from other countries, including many in Western Europe, indicates that a substantial boost in gasoline taxes would also be a powerful stimulus for innovation in automotive technologies.

Figure 7. Trends in average energy use, price and size of new US refrigerators

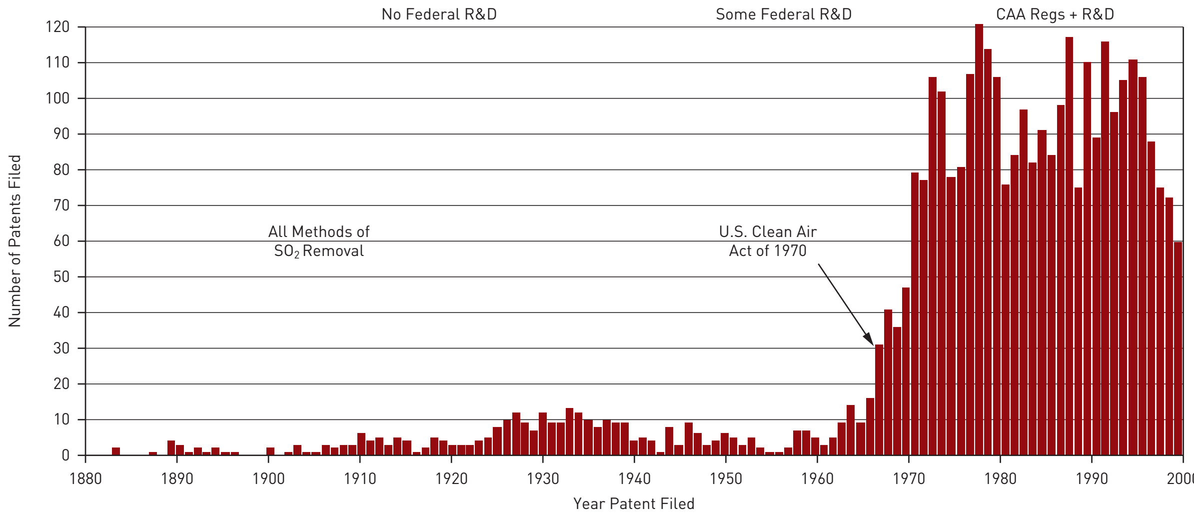


Source: Rosenfeld, 2008

Energy efficiency standards also have reduced the average energy use of major household appliances including refrigerators, dishwashers, and air conditioners. Figure 9, for example, shows the dramatic decrease in the average energy consumption of new refrigerators—then the most energy-intensive home appliance in the US—following the adoption of California state standards beginning in the 1970s, and subsequent national standards beginning in 1990. As a result of technological innovations, the average annual energy use of refrigerators was reduced to a third of its 1975 value. At the same time, the average retail price of a new refrigerator fell by a factor of two, even as the average size of new units increased. The overall savings in electricity demand avoided the need for many new power plants and their associated air pollutant and GHG emissions.

The case of sulfur dioxide (SO_2) emissions from electric power plants further illustrates the potential influence of performance standards on innovation for environmental control technologies. Stringent national limits on SO_2 emissions from new coal-fired plants

Figure 8. US patenting activity in sulfur dioxide removal technologies, 1880-2000



Source: Taylor, *et al.*, 2005

were adopted in the US starting in 1970. The result was a dramatic rise in "inventive activity" as measured by the number of US patents filed (from around the world) in the area of SO_2 control, as seen in figure 10. As post-combustion capture technology became required and more widely implemented, the capital costs of such systems fell by more than half over two decades, while operating costs also declined sharply (Taylor *et al.*, 2005, Rubin *et al.*, 2007). During this time the performance of such systems also improved considerably: in the 1970s SO_2 "scrubbers" typically captured 80% of the potential emissions. By 1990 the norm was about 90% SO_2 removal, climbing to 95% or more just five years later (Rubin, 2001). Today the best systems are up to 99% effective in capturing SO_2. If CO_2 capture and storage technologies are to become a cost-effective option for GHG reductions, similarly sustained cost and performance improvements will likely be needed (Rubin, 2009). This history of post-combustion SO_2 capture suggests that well-crafted regulatory policies can help accomplish that goal.

The regulatory policies illustrated above are examples of what are often referred to as "command-and-control" regulations that compel polluters or manufacturers to meet specified levels of technology performance at individual facilities. The more recent adoption of "market-based" regulations, such as the cap-and-trade systems adopted for compliance with acid rain legislation and summer ozone control (Yeh *et al.*, 2005), gives polluters greater flexibility in complying with national or regional requirements for an overall level of emissions reduction. Such flexibility can significantly lower the cost of compliance.

An economy-wide cap-and-trade program is a regulatory policy approach that has been widely advocated and proposed as the most cost-effective means of greenhouse gas mitigation (e.g., Jaffe *et al.*, 2003). This approach is also the centerpiece of the current Emissions Trading System (ETS) for carbon dioxide emissions in the European Union. Alternatively, many economists advocate a tax on GHG emissions as the preferred market-based approach for

> "Voluntary technology policy measures alone will not be sufficient to stabilize GHG (greenhouse gases) levels. Sufficiently stringent regulatory policies are also needed to limit GHG emissions and to foster technology innovation"

reducing GHG emissions (NRC, 2010a). Both approaches can stimulate innovation by establishing economic incentives and markets for emission-reduction measures. In the case of cap-and-trade, this requires a sufficiently stringent cap, while in the case of an emissions fee, a sufficiently stringent tax. Because there is less historical experience with such market-based regulations, there is limited empirical evidence of their effectiveness in stimulating technology innovations that reduce environmental emissions. However, in the case of SO_2 control, a study of patent data found that the US cap-and-trade program enacted in 1990 fostered innovations that lowered the cost of operating SO_2 capture units and improved their SO_2 removal efficiency (Popp, 2003). Studies also found that the SO_2 cap-and-trade program promoted changes in the internal procedures of regulated firms as well as innovations and investments by upstream suppliers (Burtraw *et al.*, 2005). Strong theoretical grounds also support a major role for market-based policies in an overall strategy for dealing with climate change.

CHOOSING POLICY OPTIONS

The merits and limitations of alternative policies for climate change mitigation is a topic widely discussed in the literature and debated in policy forums. Inevitably, the choice of policies adopted by any nation, either unilaterally or as part of an international accord, will depend on many factors and circumstances, discussion of which is beyond the scope of this paper. Rather, the preceding discussion was intended to illustrate some of the ways in which policy choices can affect technological innovation for GHG mitigation. Similarly, we note that other types of policies, such as patenting and anti-trust enforcement, can also have an indirect influence on innovation, as discussed by Alic *et al.* (2003).

In most cases, the preferred path for climate change mitigation and technology innovation will be a combination of policies that offer both "carrots" and "sticks". The simple but important message of this section is that voluntary technology policy measures alone will not be sufficient to stabilize GHG levels. Sufficiently stringent regulatory policies are also needed to limit GHG emissions and to foster technology innovation.

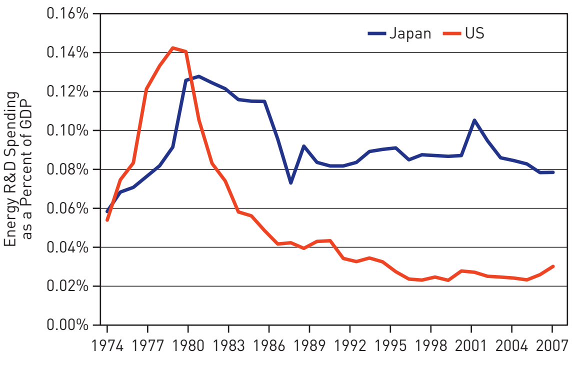
RESOURCE NEEDS FOR TECHNOLOGICAL INNOVATION

Achieving climate change goals will require not only a set of policy drivers, but also an infusion of financial and human resources to support each stage of the technological-change process depicted earlier in figure 6. Such resources are especially critical for the technology-innovation stage. In particular, there are significant needs for increased financial support for R&D and for people with the requisite training, skills and creativity to innovate—not only with regard to technologies for energy supply and demand, but also in other sectors that emit GHGs, including agriculture, forestry, and manufacturing.

The present outlook for a major infusion of such resources is decidedly mixed. In recent years, for example, China—which is now the largest emitter of GHGs in the world—has embarked on a major expansion of investment in "green" energy technologies that has propelled it to become the world's leading manufacturer of photovoltaic solar cells, as well as a dominant force in wind power systems. China is also investing heavily in nuclear power generation, and is developing a number of clean coal technologies, including carbon capture and storage systems.

In contrast, national government funding for energy R&D in the United States has declined sharply over the past three decades. In 2008, such funding was less than a fifth of what it was in 1980, in real terms. While federal energy R&D funding in the US has increased in the past few years—including a sharp rise in 2009 as part of an economic stimulus program—US expenditure for energy R&D remains much lower than for other key areas of science and technology such as space and health (NRC, 2010a). Compared to many other industrialized countries (including Canada, Denmark, Finland, France, Japan, Korea, Norway, and Sweden), the US also spends substantially less on energy R&D as a fraction of gross domestic product (GDP) (IEA, 2009). This is illustrated in figure 11, which compares government spending on energy R&D by the US and Japan as a percentage of GDP. For the

Figure 9. Government spending on energy R&D in Japan and the United States, 1974–2008



Source: NRC, 2010a

past three decades, the US percentage has been considerably lower than that of Japan. While in absolute terms the US spending is higher than in other smaller nations, the normalized data suggest that energy R&D is a lower national priority in the United States than in many other industrialized countries.

Ultimately, the private sector must play the major role in technology innovation if the climate change problem is to be dealt with successfully. Reliable data on private-sector funding of energy-related R&D is less readily available. Estimates by the International Energy Agency (IEA) and others suggest that the current rate of R&D spending by the energy industry is far below that of industries such as pharmaceuticals, biotechnology, and software and computer services—industries whose profitability depends more strongly on the ability to create new or improved products. Within the energy sector, the electric-power industry tends to have the lowest rate of R&D spending as a percentage of sales (a widely-used indicator) (NRC, 2010a). This suggests that a significant increase in private-sector investment in R&D will be needed to develop and commercialize new low-emission technologies to address climate change. In turn, government policies must provide the signals and potential markets needed to stimulate private-sector investment in R&D to reduce greenhouse gas emissions.

Technology innovations to reduce GHG emissions will also require increased numbers of skilled workers, especially engineers and scientists in a wide variety of disciplines (including the social sciences). Limited data for the US suggests that the energy industry currently has far fewer R&D workers as a percentage of the total workforce than the average for all US industries. Over the past two decades the percentage of US college graduates in engineering fields has also declined significantly (NRC, 2010a). While other countries exhibit more favorable trends, increased efforts will be needed to direct human resources and talent to focus on innovations that support climate change mitigation.

"Technological change on a massive scale will be needed over the coming decades to achieve the international goal of stabilizing atmospheric levels of greenhouse gases (GHGs) at levels that avoid dangerous impacts"

CONCLUSION

While the study of technological innovation historically has been motivated by a focus on economic development and competitiveness in a market economy, the links between innovation and the attainment of environmental quality goals has become a subject of growing interest. This paper has discussed the critical role of technology innovation in addressing the problem of global climate change—arguably the most pressing environmental challenge we presently face.

As elaborated in this paper, technological change on a massive scale will be needed over the coming decades to achieve the international goal of stabilizing atmospheric levels of greenhouse gases (GHGs) at levels that avoid dangerous impacts. This will require replacing current GHG-intensive technologies—especially energy technologies based on fossil fuels (oil, gas and coal)—with newer technologies that emit fewer or no greenhouse gases. In many cases this will require advanced technologies that have not yet been developed or adopted on a significant commercial scale, or which have not yet been invented.

Studies of technological change show that it is a complex process involving interactions among all stages of the process (invention, innovation, adoption, and diffusion of new technology into the marketplace). In general, gains from new technologies are realized only with their widespread adoption, a process that usually takes considerable time.

Government policies influence outcomes at each stage of this process. The stage of technological innovation—which leads to the development of new processes and technologies—is especially uncertain because development pathways and the likelihood of success cannot be predicted with confidence. Nor does the development of a new technology guarantee its commercial viability.

The role of government policies is especially critical in fostering innovations that address the problem of climate change. In the absence of government mandates or incentives to mitigate the problem, there are few if any markets for new technologies whose sole purpose is to reduce emissions to the environment (air, water or land). Thus, to achieve the large reductions in greenhouse gas emissions needed to reduce the risks of climate change, a broad portfolio of policies is required—not only to foster technological innovation, but also the subsequent adoption of new technologies by a large range of actors including individuals, governments, and firms of all size.

The policy portfolio to foster innovation should include a combination of "sticks" in the form of regulatory policies that directly or indirectly set limits on GHG emissions (such as through market-based mechanisms, technology performance standards, or a combination of measures), together with "carrots" in the form of technology policies that provide voluntary incentives to encourage technology innovation and deployment (such as through support for R&D, tax credits, loan guarantees, government procurement programs and other measures). To realize the full benefits of technological innovation, the policy portfolio also should support diffusion of knowledge, such as through financial support for education and training, along with other measures.

Although R&D alone is not sufficient to achieve widespread technological change, it is nonetheless a critical element of the policy portfolio needed to foster innovations that reduce GHG emissions. As discussed in this paper, substantial increases in government support for energy-related R&D are required to address the challenges of climate change. Large increases are also needed in private-sector support for R&D, especially in energy-related industries. Government policies again play a vital role by establishing the requirements and market signals needed by the private sector to justify R&D investments.

Finally, reducing GHG emissions through innovations in technology and institutions will require increased numbers of skilled workers, especially engineers and scientists across a wide variety of disciplines, including the social sciences. At the end of the day, it is people who innovate. Both government and the private sector have critical roles to play in attracting and retaining the best and brightest people worldwide to address the challenges and invent the opportunities for mitigating global climate change.

REFERENCES

BURTRAW *et al.* (2005), "Economics of Pollution Trading for SO_2 and NO_x", *Annual Review of Environment and Resources* 30, pp. 253-289, Annual Reviews, Palo Alto, CA.

CATF, Consortium for Science, Policy & Outcomes and Clean Air Task Force (2009), *Innovation Policy for Climate Change: A Report to the Nation, A Joint Project of the Consortium for Science, Policy & Outcomes and Clean Air Task Force,* Clean Air Task Force, Boston, MA. September

CLARKE, L. *et al.* (2009), "International Climate Policy Architectures: Overview of the EMF 22 International Scenarios", *Energy Economics* 31 (Supplement 2): pp. 64-81.

FAWCETT, A. A., *et al.* (2009), "Overview of EMF 22 US Transition Scenarios", *Energy Economics* 31 (Supplement 2), pp. 198-211.

IEA, International Energy Agency (2000), *Experience Curves for Energy Technology Policy,* Paris, France.

IEA, International Energy Agency (2009), *Ensuring Green Growth in a Time of Economic Crisis The Role of Energy Technology*, Paris, France, p. 7.

IEA, International Energy Agency (2010), *Annual Energy Review 2009,* Washington D.C.: Energy Information Administration, US Department of Energy.

IPCC (2007a), *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* S. Solomon et al., eds. Cambridge University Press, Cambridge.

IPCC (2007b), *Climate Change 2007: Mitigation, Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, B. Metz *et al.* (eds.), Cambridge: Cambridge University Press.

JAFFE, A. B., R. G. NEWELL and R. N. STAVINS (2003), "Technology Policy for Energy and the Environment", in A. Jaffe, J. Lerner, and S. Stern (eds.), *Innovation Policy and the Economy*, Conference Report, National Bureau of Economic Research.

KYLE, P. *et al.* (2009), "The Value of Advanced Technology in Meeting 2050 Greenhouse Gas Emissions Targets in the United States", *Energy Economics* 31(2), pp. 254-267.

LEE, J. *et al.* (2010), "Forcing technological change: A case of automobile emissions control technology development in the US", *Technovation* 30, pp. 249–264.

MCDONALD, A. and L. SCHRATTENHOLZER (2002), "Learning Curves and Technology Assessment", *International Journal of Technology Management* 23, pp. 718-745.

NHTSA, National Highway and Traffic Safety Administration (2010), "NHTSA and EPA establish new national program to improve fuel economy and reduce greenhouse gas emissions for passenger cars and light trucks", at: http://www.nhtsa.gov/fuel-economy, Accessed August 15.

NRC, National Research Council (2010a), *Limiting the Magnitude of Climate Change,* Washington DC: The National Academies Press.

NRC, National Research Council (2010b), *Advancing the Science of Climate Change,* Washington, DC: The National Academies Press.

NRC, National Research Council (2010c), *Adapting to the Impacts of Climate Change,* Washington, DC: The National Academies Press.

POPP, D. (2003), "Pollution Control Innovations and the Clean Air Act of 1990", *Journal of Policy Analysis and Management* 22(4), pp. 641-660.

ROSENFELD, A. H. (2008), "Energy Efficiency: The first and most profitable way to delay Climate Change", Oppenheim Lecture, Los Angeles: UCLA Institute of the Environment, University of California, February 25, at: http://www.energy.ca.gov/commission/commissioners/rosenfeld.html. Accessed March 15.

RUBIN, E. S. (2001), *Introduction to Engineering and the Environment,* Boston, MA: McGraw-Hill.

RUBIN, E. S. *et al.* (2004), "Learning Curves for Environmental Technology and their Importance for Climate Policy Analysis", *Energy* 29, pp.1551-1559.

RUBIN, E. S., *et al.* (2007), "Use of Experience Curves to Estimate the Future Cost of Power Plants with CO_2 Capture", *International Journal of Greenhouse Gas Control* 1, pp. 188-197.

RUBIN, E. S. (2009), *A Performance Standards Approach to Reducing CO_2 Emissions from Electric Power Plants,* Arlington, VA: Coal Initiative Reports, White Paper Series, Pew Center on Global Climate Change, May.

TAYLOR, M. R., E. S. Rubin and D. A. Hounsell (2005), "Control of SO_2 Emissions from Power Plants: A Case of Induced Technological Innovation in the US", *Technological Forecasting and Social Change* 72(6), pp. 697-718, July.

YEH, S. *et al.* (2005), "Technology Innovations and Experience Curves for Nitrogen Oxides Control Technologies", *Journal of the Air & Waste Management Association* 55: December, pp. 1827-1838.

Life Innovation with Therapeutic Robot, Paro

Takanori Shibata
Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology. Japan

INTRODUCTION

Interaction with animals has long been known to be emotionally beneficial to people. In recent years, the effects of animals on humans have been researched and proved scientifically. Friedmann investigated the one-year survival rate of patients who were discharged from a coronary-care unit and found that survival among those who kept pets was higher than among those who did not (Friedmann *et al.*, 1980). Baun *et al.* reported that patients' blood pressure dropped when they petted their dogs (Baun *et al.*, 1984). Garrity *et al.* studied older people who were socially isolated and had lost their partner during the previous year, and found that the intensity of depression among those who had no pets was higher than among those who did (Garrity *et al.*, 1989). Lago *et al.* researched through telephone interviews the influence on older people of owning pets. They found that mortality and attrition were higher for former owners than for current owners (Lago *et al.*, 1989). Hart *et al.* studied the social influence of animals on people, and found that those with dogs were approached in a friendly way by strangers more frequently than those without dogs (Hart *et al.*, 1987).

In medical applications, especially in the United States, animal-assisted therapy and activities (AAT&AAA) are becoming widely used in hospitals and nursing homes (Delta Society, 1996; Fine, 2006). AAT has clear goals set out in therapy programs designed by doctors, nurses and social workers, in cooperation with volunteers, while AAA refers to patients interacting with animals without particular therapeutic goals, and depends on volunteers. AAT and AAA are expected to have three effects:

1. a psychological effect (relaxation, motivation);
2. a physiological effect (the improvement of vital signs);
3. a social effect (stimulation of communication among inpatients and caregivers).

For example, a hospitalized child, who was in significant pain on account of his illness, was afraid to get up and walk around. However, when he was asked to take a therapy dog for a walk, he immediately agreed and walked off happily, as if all his pain had disappeared. Moreover, the dog acted as a medium for interaction between him and the other children (Kale, 1992). In another case, a boy who was born as a crack-exposed baby was unable to speak or walk. However, through interaction with therapy dogs and birds, both his linguistic and motor abilities improved (Delta Society, 1991).

For AIDS patients, it is important to reduce their stress as it greatly affects the complications of immune deficiency. AAT helps them to relax and stay connected with the world (Haladay, 1989).

In addition, AAT and AAA at nursing homes help to rehabilitate older people, make them laugh, and bring happiness to patients who have only a short time to live (Gammonley and Yates, 1991). AAT reduces loneliness in residents of long-term care facilities (Banks and Banks, 2002). The presence of therapy animals has been particularly useful in reducing agitated behavior, decreasing episodes of verbal aggression and anxiety, and increasing social interaction among institutionalized old people suffering from dementia (Richeson, 2003; Fick, 1993; Fritz *et al.*, 1995).

However, most hospitals and nursing homes, especially in Japan, do not allow animals, although they recognize the positive effects of AAT and AAA. They are afraid of the negative impact of animals on human beings, in the form of allergic reactions, infection, bites, and scratches.

Society in most advanced countries is aging. Among them, Japan has the most old people: 23.1 percent of the population (29.4 million out of 127 million) in 2010 are more than 65 years old. Many old people stay healthy, but some of them need care at home or in institutions, depending on their levels of physical disabilities and cognitive disorders. Among the cognitive disorders, dementia presents the most problems. About 2 million people suffer from dementia in Japan In 2010.

Mental-health care for the elderly is an important issue for caregivers at nursing homes (Daies and Knapp, 1981) as depressive disorders are common among old people who have experienced difficult situations such as the loss of their family, friends, social roles, and physical functions (WPA/PID). In addition, those who suffer from mental diseases such as dementia, hallucinations, aggression and wandering cause psychiatric and behavioral disturbances (ADI, 1999). Such disorders have a negative impact on their quality of life and that of their caregivers. In order to improve communication with older people, caregivers conduct recreational activities, such as singing songs, coloring, drawing pictures, and origami. However, some people are too embarrassed to sing songs, and others lack finger mobility when attempting to draw. In addition, communication is complicated on account of a lack of common topics.

This paper introduces robot therapy for "life innovation", and discusses its potential in care for the elderly. The next section describes the new area of robotics: human-interactive robots for psychological enrichment. We then explain the required functions for therapeutic robots and the seal robot, Paro. Finally, we describe examples of robot therapy for older people, and present our conclusions.

HUMAN-INTERACTIVE ROBOTS FOR PSYCHOLOGICAL ENRICHMENT

Industrial robots have been used widely in manufacturing industries since the early 1960s. Industrial robots typically perform welding, assembling, painting, packaging, and palletizing in automotive manufacturing and other industries. Such robots work very fast and accurately, although initially they need to be taught by a human operator and their environment needs to be especially designed to enable them to accomplish their tasks. Most industrial robots are considered a potential danger to humans, and therefore are kept isolated from them.

Meanwhile, the rapid development in high technology has produced robots not only for factories but also for our living environment, such as homes, offices, and hospitals. For example, wheelchair robots enable elderly people to move easily outside (Matsumoto *et al.*, 2006). Robotic suits, which can expand the physical capability of humans, are expected to reduce the workload of caregivers (see HAL). A horseback-riding robot promotes a patient's physical strength (see JOBA). Human-interactive robots for psychological enrichment, in particular, are expected to be a new application of robotics and are attracting many researchers and companies (Shibata, 2004). Human-interactive robots are designed for entertainment, communication (social activity), guidance, education, welfare, mental therapy, and other purposes. Various types of robots, such as humanoid, animal, and robots with a unique appearance, have been developed.

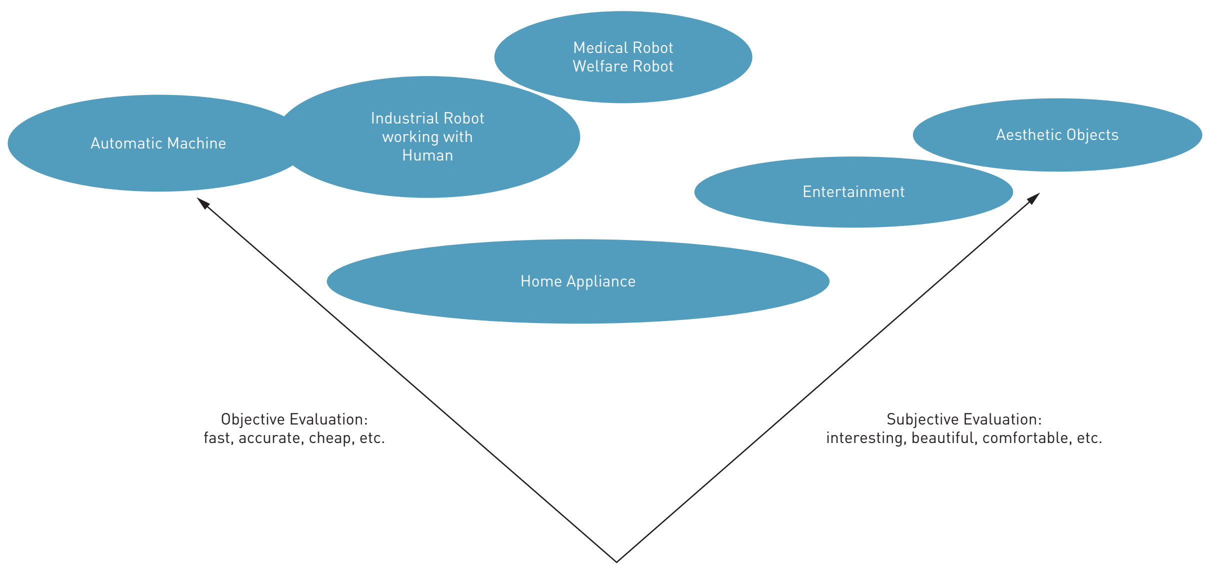
These robots offer more interaction with humans than industrial robots. They are evaluated not only in terms of objective measures, such as speed and accuracy, but also in terms of subjective measures regarding interacting with humans, such as providing comfort and happiness. Robots for entertainment are good examples of the importance of a subjective evaluation of their value (fig. 1).

There are four categories of human interactive robots for psychological enrichment, in terms of their relationship with humans: 1. performance robots; 2. tele-operated performance robots; 3. operation, building, programming, and control robots; and 4. interactive autonomous robots.

1. Performance robots:

Performance robots have a long history and they execute movements that express meanings to humans, mostly for entertainment. Mechanical puppets that were able to play an organ, draw pictures, and write letters were developed in Switzerland in the 18th century. Karakuri dolls were

Figure 1. Objective and Subjective Measures for Evaluating Artifacts



developed in Japan during the same era to dance, perform magic tricks, and so on. Recently, many performance robots have been used at exhibitions, museums, movies, and amusement parks such as Disneyland and Universal Studios. Recent humanoid robots, such as Honda's ASIMO and Sony's QRIO, can be included in this category (Hirai, 1998; Kuroki *et al.*, 2002). Performance robots are able to amuse sizable audiences at any time. However, their movements are probably preprogrammed and mostly repetitive; therefore, they are not usually very interactive with humans. A high degree of complexity in performance robots is required in order to keep humans amused.

2. *Tele-operated performance robots:*

Tele-operated performance robots are controlled remotely by hidden operators. Their movements can appear reactive to the humans who interact with them because the operator, based on the audience's actions, sends commands to the robots to simulate reactive behavior. At exhibitions or amusement parks, for example, human-type robots are used as tele-operated performance robots.

3. *Operating, building, programming, and controlling robots:*

Humans derive a great deal of entertainment from operating, building, programming, and controlling robots. Moreover, they can watch the performance of the robots that they are operating. A simple example of this is the "UFO catcher", a stuffed-animal game machine at amusement centers. Building and programming a robot is also included in this category. Contests between robots such as Micro-mouse, RoboCup (robot football, Kitano *et al.*, 1998), and RoboOne (robot wrestling, see Robo-One) are popular examples, as are LEGO-Mindstorms and I-Blocks. Because building and programming robots can stimulate children's creativity, this activity combines entertainment with education, and is often referred to as "edutainment" (Druin and Hendler, 2000; Lund, 2004).

4. *Interactive autonomous robots:*

Interactive autonomous robots connect with humans in the physical world. They use verbal and nonverbal communications, depending on their functions. Contrary to the interactions of robots in the other categories, human-robot interactions are mostly personal. For example, Sony's dog robot, AIBO, which is designed for entertainment, has a mechanical appearance and attracts people's interest using nonverbal communication (Fujita, 2004). The communication robot, Ifbot, produces conversation using facial expressions and a large number of prepared conversation scenes (see ifbot). The human-friendly information terminal, PaPeRo, can control home electric appliances and collect information via the Internet by voice command, and even entertain people by dancing and playing games (see PaPeRo). Guide robots in museums and exhibitions (Bischoff and Graefe, 2004), and mental-commitment robots described in this paper also belong to this category.

In the area of welfare and mental therapy, Shibata *et al.* studied and developed a mental-commitment robot, which aims to engender mental effects, such as pleasure and relaxation, in its role as a personal robot (see Shibata, 1996-2009). They also proposed "robot therapy," which used robots as substitutes for animals in animal-assisted therapy and activity. Robot therapy targets the people in medical and welfare institutions where animals are not allowed. A seal-type mental-commitment robot, named Paro (fig. 2), was developed especially for robot therapy and is used at pediatric hospitals and facilities for elderly people in several countries. Recent research has revealed that robot therapy has the same effects on people as animal therapy. Robot therapy, in particular, is recognized as a new method of mental health care for elderly people (including dementia patients).

THERAPEUTIC ROBOT

Required Functions

In robot therapy, it is important to stimulate people's knowledge and experience of animals through interaction with the robots, and to touch their feelings when they are interacting with animals. Therefore, the shapes, feelings of touch, autonomous behaviors and responses that mimic animals are the features that are required in the robots.

In addition, the devices are used not only in people's homes but also in hospitals and nursing homes. Many people who have lost their physical strength and healing capability on account of aging or illness are expected to interact with the robots. Therefore, the robots must be easily accepted by people and also be harmless and hygienic. A cause of concern regarding such robots is that individuals are expected to physically interact with them by touching and hugging them, and therefore there exists a possibility of them being harmed. Furthermore, in case robots are used by hospital patients with deteriorated immune systems such as leukemia, the robots may transmit germs. Furthermore, some people visit medical-welfare facilities for a few hours at a time for day care and ambulant treatment, but some stay or are hospitalized for years (e.g., in nursing homes for long-term care). Therefore, the robots have to sustain long-term interaction with people in their daily lives.

These robots are used by doctors, nurses, therapists, caregivers, and volunteers for long periods of time. In addition, users play with them whenever they want. Therefore, it is important that the robots be designed in

Figure 2. Seal Robot, Paro

such a manner that anyone is able to operate them, and that no specialized knowledge is required to do so.

Mental-commitment robot, Paro:

Mental-commitment robots are not intended to perform physical work or service (see Shibata, 1996-2009). Their function as personal robots is to engender mental effects, such as pleasure and relaxation. These robots have a purpose and act independently, although, like living organisms, they receive stimulation from the environment. Their actions during interaction with people make it seem that they have hearts and feelings.

Mental-commitment robots are able to stimulate the different senses of human beings through physical interaction. Thus, the primary characteristic of mental-commitment robots is nonverbal communication. A basic psychological experiment was conducted on the subjective interpretation and evaluation of robot behavior following human-robot interaction. This study showed the importance of appropriately stimulating the human senses and extracting associations. Sensor systems, consisting of visual, aural, and tactile senses for robots, were studied and developed. A plane soft tactile sensor was developed to cover the robot to enhance bodily contact between people and the robot (Shibata, 2004). This sensor can detect position and force when people touch the robot, and at the same time, it is soft to touch.

There are three categories of animal robots:

1. familiar animals (e.g., dog, cat);
2. unfamiliar animals (e.g., seal);
3. imaginary animals or characters.

The dog, cat, and seal robots were developed as models. Each robot operates using its three internal elements, sensory information from its sensors, and its own diurnal rhythm (morning, afternoon, and night) to perform various activities during its interaction with people.

Subjective evaluations of the cat and seal robots were conducted using a questionnaire (Shibata *et al.*, 1999; Shibata and Tanie, 2000). Both robots were evaluated highly. However, subjects complained about the softness and reactions of the cat robot in comparison with their knowledge of real cats. On the other hand, most people do not have much knowledge of seals, and hence were unable to compare the seal robot with what they knew about them. Therefore the evaluation of the seal robot following the interaction was higher. These results revealed that more people accepted the unfamiliar animal shape.

Cross-cultural studies on the subjective evaluation of the seal robot were conducted in seven different countries: Japan, the UK, Sweden, Italy, Korea, Brunei, and the US (Shibata *et al.*, 2002; Shibata *et al.*, 2009). The data were obtained from about 2,000 respondents. The subjective evaluation provided overall high scores, and revealed that the seal robot could be widely accepted despite cultural and religious differences. However, from the results of the principal component analysis, there were two different sets of expectations regarding Paro: one was as a pet, and the other was for therapy (fig. 3). In Asian countries, Japan and Korea, people accepted Paro as a pet, but not for therapy. In European countries, the UK, Sweden, and Italy, people accepted Paro for therapy, but not as a pet. In the US and Brunei, people accepted Paro as a pet and for therapy. This may be because of the different relationships between humans and pets, and the attitudes toward robots in general. In Asia, pets are popular, but their status is lower than that of humans. In addition, animal therapy is not known or trusted very much. In Western countries, however, pets are loved and their status is equal or higher than that of humans. In addition, animal therapy is well known and is practiced. However, robots are considered to be dangerous or even evil, and there is a fear that jobs might be taken by them.

The seal-type mental-commitment robot, Paro, was designed both as a pet and for therapy. Even though Paro's surface is covered with pure white or light gold fur (fig. 2), Paro contains high-tech mechanisms (fig. 4). Ubiquitous surface tactile sensors are inserted between the hard inner skeleton and the fur to create a soft, natural feel and to permit the measurement of human contact with Paro. His weight is approximately 2.7 kg., his weight and size simulating those of a human baby. Paro has four senses: sight (light sensor), audition (determination of sound source direction and speech recognition), balance, and a tactile sense. In addition, Paro has a temperature sensor, which maintains a constant warm body temperature. His mobile parts are as follows: vertical and horizontal neck movements, front and rear paddle movements, and independent movement of each eyelid, which is important for creating facial expressions.

Figure 3. Expectations of Paro from Results of Questionnaires in Seven Countries

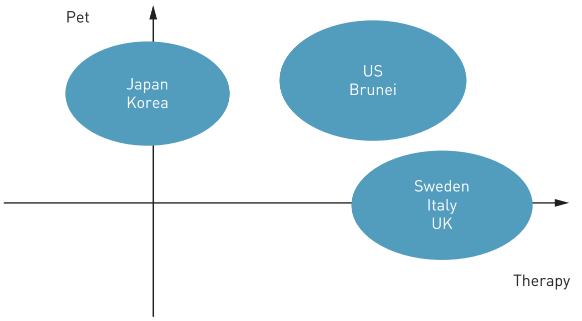


Figure 4. Sensors, Actuators, Distributed Control System, and Artificial Intelligence in Paro

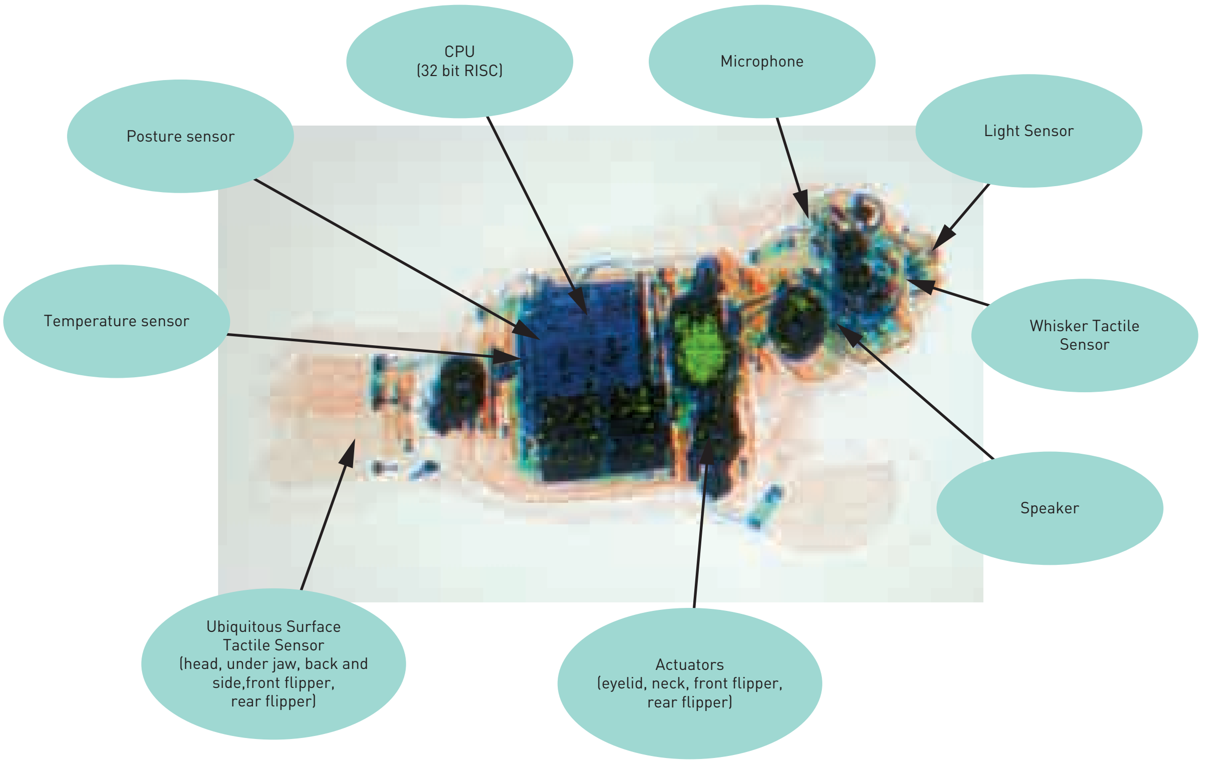


Figure 5. Handcrafted Processes

(a) Process of Trimming Fur

(b) Process of Sewing Eyelashes

In order to be evaluated highly by people interacting subjectively with him, Paro was functionally designed to be soft and evoke a feeling of warmth. Each Paro was trimmed with artificial fur, and eyelashes were sewn onto him by craft workers to achieve high quality (fig. 5). In addition, his artificial fur is antibacterial and dirt resistant, and will not fall out. An electromagnetic shield was installed in the internal circuit to prevent any ill effects on pacemakers. The withstand-voltage test, a drop test, a one hundred-thousand times stroking test, and a long-term seven- year clinical test confirmed that Paro is very safe and durable. Paro was designed to be simple enough for anyone to operate. He has only one on/off switch for power, and a pacifier-type charger. Artificial intelligence enables users to learn Paro's name and become acquainted with his behavior, thereby preventing them from losing interest, allowing them to gradually build a relationship with him, and showing their affection for him. In addition, baby harp seals were investigated so that their liveliness and cuteness could serve as models for the robot; real baby seal calls were sampled and used.

Robot therapy

Robot therapy using seal robots is conducted at hospitals and nursing homes in many countries: Japan, Sweden, Denmark, Italy, the US, and so on. Robot therapy consists of robot-assisted therapy programs designed by doctors, nurses, and social workers, and also robot-assisted activity, which allows patients to interact with robots without any particular therapeutic goals. Such activities do not depend on volunteers, but are conducted by facility staff. Robot-assisted therapy is mainly conducted at medical facilities, such as hospitals and clinics, and also at welfare facilities such as nursing homes (fig. 6)

Robot therapy for the elderly

As an example of robot therapy, Paro was used to assist elderly people at a day service center (Shibata *et al.*, 2001; Saito *et al.*, 2002; Wada *et al.*, 2004). To investigate the effects of seal robots on the elderly, their moods were evaluated using face scales (Lorish and Maisiak, 1986) and questionnaires. Changes in their reactions to stress were measured by the hormones in their urine: 17-ketosteroid sulfate (17-KS-S) and 17 hydroxycorticosteroids (17-OHCS) (Selye, 1970; Nishikaze *et al.*, 1995). In addition, the stress that the nursing staff experienced was investigated by questionnaires -- i.e., the burnout scale (Pines, 1981). The day service center was provided with seal robots for five weeks, and the feelings of the old people improved as a result of their interaction with the robots. Urine samples showed that their ability to overcome stress also improved. Moreover, the stress levels of the nursing staff decreased because the old people required less supervision while interacting with the robots.

A long-term experiment was conducted at a health-service facility, starting in August 2003 (Wada *et al.*, 2004). Approximately 10 people joined the interaction with Paro for one hour, twice a week. One or two caregivers managed the interaction with Paro. To investigate the effects of Paro, face scales and geriatric depression scales were used (Yesavage, 1988). The results showed that the feelings of the old people improved over the year, and depression in the participants was also reported to have decreased. The caregivers commented that interaction with Paro made the old people laugh and become more active. Their facial expressions changed, softened, and brightened. On the day of the activity, they looked forward to interacting with Paro, sitting down and waiting even before the interaction session began. Some people who usually stayed in their rooms came out and willingly joined the activity. In addition, Paro encouraged people to communicate with each other as well as with the caregivers by becoming their common

Figure 6. Interactions between Older People and Paro

topic of conversation. Thus, the general atmosphere became brighter. Even now, these older people enjoy playing with Paro.

In another example, Paro was introduced in the public area of a care house, a type of communal housing for older people, and was activated for over nine hours each day for the researchers to investigate the effects of free interaction with him (Wada and Shibata, 2007).To examine the psychological and social effects, each subject was interviewed and his or her social network analyzed. In addition, the activities of the residents in public areas were video recorded. For physiological analysis, residents' hormones in urine; 17-KS-S and 17-OHCS, were analyzed. The results indicate that interaction with Paro increased their social interaction. Furthermore, the urine tests showed that the reactions to stress of the subjects' vital organs improved after interacting with Paro.

Effects on patient with dementia

Dementia is a major problem in care for the elderly. According to Alzheimer's Disease International (ADI), an estimated 24.4 million people suffer from dementia worldwide, and the number will increase to 82 million by 2040. Dementia is a progressive, disabling neurological condition that occurs in a wide variety of diseases. The most common cause of dementia is Alzheimer's disease (AD), which accounts for approximately half of the people with dementia. Other causes include vascular disease, Lewy body dementia, and many other diseases (see ADI, 1999). Psychiatric and behavioral disturbances, such as personality changes, hallucinations, paranoid ideas, aggression, wandering, and incontinence are common features of dementia and are the leading causes of the need for long-term care (Garrity et al., 1989). Donepezil, physical exercise, and diet cure are expected to slow the progress of dementia (Andrade and Radhakrishnan, 2009). But unfortunately, there is no permanent cure for dementia at present. Recent data suggest that art, music, and learning, which stimulate patients' emotions and brain, can slow its progression once it has begun (Kimura *et al.*, 2002; Brotons and Koger, 2000; Kawashima, 2002). However, there is room for improvement in all such treatments.

As for the interaction between Paro and dementia patients in nursing homes, behavioral improvements were observed in several cases. For example, a patient who moaned continuously was able to relax and then started to talk to the therapist (Marti *et al.*, 2006). After playing with Paro, another patient who often tried to return home stopped doing so; her wandering symptom improved.

Robot therapy for dementia patients was conducted at a neurosurgery clinic to investigate the physiological influences of the treatment (Wada *et al.*, 2008). Diagnosis Method of Neuronal Dysfunction (DIMENSION) was used to record each patient's EEG before and after 20 minutes of robot therapy (Musha *et al.,* 2002). In addition, a questionnaire concerning each subject's impression of Paro was conducted. The results showed that 50% of 14 valid subjects' condition of cortical neurons activity improved by interacting with Paro. This was especially true for patients who were particularly fond of Paro.

In Japan, the cost of care for a dementia patient by social welfare insurance is about 40,000 USD per year, and the patient's life expectancy is about 8 years. In Denmark, the cost is about double. This represents an enormous burden on the municipalities that provide long-term care insurance. Useful and convenient methods for the prevention of dementia are urgently needed. Paro has been on the market since 2005. The price is about 4600 euros. However, the running costs consist almost only of a battery-recharging fee as the battery is designed to last for over 10 years. Paro can be used even without a specially-trained therapist although people must attend a one-day seminar in the EU in order to use him. Several municipalities in Japan know about Paro's effects and support his introduction. For example, Nanto city, Toyama, bought eight Paros and introduced them to day-service centers in the city. Tsukuba city, Ibaraki, established a subsidy for purchasing Paro.

Meanwhile, a dementia-care center in Copenhagen, Denmark, investigated the effects of robot therapy on dementia patients as part of the national project "Be-Safe". Twelve Paros were introduced in 10 different places. The results obtained from the seven-month clinical trial showed that Paro had positive effects on the patients. Based on the results, the Danish government decided to introduce 1,000 units of Paro to nearly all the elderly-care facilities in Denmark. So far, more than 100 institutions, with more than 200 licensed staff, have been using Paro in 2010. In addition, other European countries including Norway, the Netherlands, and Germany, have adopted the same system, and have started to use Paro with licensed staff.

Other research

Robot therapy, using commercially-available animal-type robots, such as AIBO and NeCoRo has been attempted (Libin and Libin, 2004; Kanamori *et al.*, 2002; Tamura *et al.*, 2004). For example, Libin introduced NeCoRo to a nursing home and observed the patients' interaction (Libin and Libin, 2004). Kanamori *et al.* examined the effects of AIBO on the elderly in a nursing home (Kanamori *et al.*, 2002). By measuring the hormones in saliva, they found that stress decreased after a one-hour interaction with AIBO, and that loneliness was reduced after 20 sessions over a seven-week period. Tamura *et al.* compared the exposure of patients to AIBO with the effect of exposure to a toy dog (Tamura *et al.*, 2004). They found that AIBO did not encourage much interaction, and required more intervention from an occupational therapist.

Because they are not designed for therapy, these commercially-produced robots easily break while interacting with people. Therefore, it is difficult to use them in long-term situations.

CONCLUSION

Various robots have been developed and are being introduced into our lives as commercial products. Each robot is designed for a specific purpose. The seal-type mental-commitment robot, Paro, whose goal is to enrich daily life and heal human minds as "life innovation", is designed to maintain long-term interaction with people and provide them with psychological, physiological, and social benefits.

Robot therapy, or mental-health care using animal-type robots, is an emerging field. The results of exploratory experiments showed that Paro has a great potential to provide mental-health care for older people. However, more subjects and a control group are necessary to scientifically verify its effects. Further experiments will be conducted in this respect.

As for ethical issues, the experiments using Paro mentioned above were conducted under the supervision of the ethical committee of each organization. Only people who, and whose relatives who, agreed with robot therapy, participated. In addition, some were worried that old people would be left alone with Paro when the caregiver left them to interact with him. But, the opposite has proved true. Paro acts as a social mediator, providing a common topic for the elderly and encouraging them to communicate with each other (fig. 7). Currently, the methods used in robot-therapy are the responsibility of the caregivers and its effects are influenced by them. To ensure Paro's potential by developing effective ways to use it is the next step.

At present, approximately 1700 units of Paro have been sold worldwide (about 1500 units in Japan, 120 units in Denmark, and 100 units in other countries). Paro has a high rate of acceptance and similar psychological effects are seen in each country. However, there are initially some cultural differences in the acceptance of Paro. People in Europe tend to note the effects of interaction with Paro and accept him as a therapeutic tool because animal therapy has been used widely in hospitals and nursing homes. Therefore, all Paros are used at institutions in Europe. Meanwhile, in Asian countries, animal therapy is not commonly accepted although many people have owned pets recently. In Japan, nearly 70% of users are individuals. They tend to accept Paro as their companion more than as a therapeutic tool. In the U.S., they tend to accept Paro as a therapeutic tool as well as a companion (Shibata *et al.*, 2009), though the Food and Drug Administration (FDA) of the U.S. certified Paro as a "medical device" in 2009. Therefore, it is important to introduce Paro in a suitable manner based on the cultural differences.

Robot therapy in medical and welfare facilities is spreading in our society. Further studies should be conducted to integrate robot therapy into our societies at large.

Figure 7. Paro as Social Mediator between the Elderly and Caregiver (at an Assisted-Living Center in Denmark)

BIBLIOGRAPHY

ADI Factsheet (1999), "The Prevalence of Dementia", *Alzheimer's Disease International*, 3.

ADI Factsheet (1999), "Psychiatric and Behavioural Disturbances in Dementia", *Alzheimer's Disease International*, 7.

Andrade, C. and R. Radhakrishnan (2009), "The Prevention and Treatment of Cognitive Decline and Dementia: An Overview of Recent Research on Experimental Treatments", *Indian J Psychiatry* 51(1), pp. 12-25.

Baun, M., N. Bergstrom, N. Langston and L. Thoma (1984), "Physiological Effects of Human/Companion Animal Bonding", *Nurs. Res.* 33(3), pp. 126-129.

Bischoff, R. and V. Graefe (2004), "HERMES—a Versatile Personal Robotic Assistant", *Proceed IEEE* 92(11), pp.1759-1779.

Brotons, M. and S. Koger (2000), "The Impact of Music Therapy on Language Functioning in Dementia", *J Music Ther* 37(3), pp. 183-195.

Daies, B. and M. Knapp (eds.) (1981), *Old People's Homes and the Production of Welfare*, Routledge & Kegan Paul PLC.

Delta Society (1991), Animal-Assisted "Therapy and Crack Babies: a New Frontier", *Newsletter* 1(2).

Delta Society (1996), *Standards of Practice for Animal-Assisted Activities and Therapy*.

Druin, A. and J. Hendler (eds.) (2000), *Robots for Kids; Exploring New Technologies for Learning*, Morgan Kaufmann Pub.

Fick, K. M. (1993), "The Influence of an Animal on Social Interactions of Nursing Home Residents in a Group Setting", *Am J Occupat Ther* 47(6), pp. 529-534.

Fine, A. H. (2006), *Handbook on Animal Assisted Therapy: Theoretical Foundations and Guidelines for Practice*, 2nd edition, Academic Press.

Friedmann, E., A. Katcher, J. Lynch and S. Thomas (1980), "Animal Companions and One-year Survival of Patients after Discharge from a Coronary Care Unit", *Public Health Rep*; 95(4): pp. 307-312.

Fritz, C. L. *et al.* (1995), "Association with Companion Animals and the Expression of Non-cognitive Symptoms in Alzheimer's Patients", *J Nerv Ment Dis* 183(7), pp. 459-463.

Fujita, M. (2004), "On Activating Human Communications With Pet-Type Robot AIBO", *Proceed IEEE* 92(11), pp. 1804-1813.

Gammonley, J., J. Yates (1991), "Pet Projects Animal Assisted Therapy in Nursing Homes", *J Gerontol Nurs* 17(1), pp. 12-15.

Garrity, T., L. Stallones, M. Marx and T. Johnson (1989), "Pet Ownership and Attachment as Supportive Factors in the Health of the Elderly", *Anthrozoos* 3(1), pp. 35-44.

HAL, http://wwwcyberdynejp.

Haladay, J. (1989), "Animal Assisted Therapy for PWAs—Bringing a Sense of Connection", *AIDS Patient Care*, pp. 38-39.

HART, L., B. HART, B. BERGIN (1987), "Socializing Effects of Service Dogs for People with Disabilities", *Anthrozoos* 1(1), pp. 41-44.
HIRAI, K. (1998), "Humanoid Robot and Its Applications", *Proceedings IARP Int. Conf. Humanoid and Human Friend Robot* 1, pp. 1-4.
ifbot. http://wwwbusiness-designcojp.
JOBA, http://panasonicjp/fitness.
KALE, M. (1992), "Kids & Animals", *Interactions* 10(3), pp. 17-21.
KANAMORI, M., M. SUZUKI and M. TANAKA, "Maintenance and Improvement of Quality of Life among Elderly Patients Using a Pet-Type Robot", *Nippon Ronen Igakkai Zasshi* 39(2), pp. 214-218.
KAWASHIMA, R. (2002). "[Cognitive Rehabilitation for Alzheimer Disease -- the Learning Therapy]", *Rinsho Shinkeigaku*, 45(11), pp. 864-866.
KIMURA, S., K. KANEKO, K. NISHIDA, K. SEKINE, T. MUSHA and M. OH (2002), "[Effectiveness of the Creative Art Therapy by Using the Diagnosis Method of Neuronal Dysfunction (DIMENSION)]", *Report Kansei Fukushi Res. Cent.* 3, pp. 195-200.
KITANO, H. *et al.* (1998), "The Robocup '97 Synthetic Agents Challenge"; in *RoboCup-97: Robot Soccer World Cup I*. Springer, pp. 62-73.
KUROKI, Y., T. ISHIDA, J. YAMAGUCHI, M. FUJITA and T. DOI (2002), "A Small Biped Entertainment Robot, J. Robot", *Mechatron* 14(1), pp. 7-11.
LAGO, D., M. DELANEY, M. MILLER and C. GRILL (1989), "Companion Animals, Attitudes toward Pets, and Health Outcomes among the Elderly: A Long-Term Follow-up", *Anthrozoos* 3(1), pp. 25-34.
LIBIN, A. and E. LIBIN (2004), "Person-Robot Interactions from the Robopsychologists' Point of View: The Robotic Psychology and Robotherapy Approach", *Proceed. IEEE* 92(11), pp. 1789-1803.
LORISH, C. D. and R. MAISIAK (1986), "The Face Scale: A Brief, Nonverbal Method for Assessing Patient Mood", *Arthritis and Rheumatism* 29(7), pp. 906-909.
LUND, H. (2004), "Modern Artificial Intelligence for Human-Robot Interaction", *Proceed. IEEE* 92(11), pp. 1821-1838.
MARIAN, R. Banks and William A. BANKS (2002), "The Effects of Animal-Assisted Therapy on Loneliness in an Elderly Population in Long-Term Care Facilities", *J Gerontol A: Biol Sci Med Sci* 57, pp. 428-432.
MARTI, P., M. BACIGALUPO, L. GIUSTI, C. MENNECOZZI and T. SHIBATA (2006), "Socially Assistive Robotics in the Treatment of Behavioural and Psychological Symptoms of Dementia", *Proceed. Int. Conf. BioRob*, pp. 483-488.
MARTI, P., V. PALMA, A. POLLINI, A. RULLO and T. SHIBATA (2005) "My Gym Robot", *Proceed. Symp. Robot Companions,* pp. 64-73.
MATSUMOTO, O., K. KOMORIYA, T. HATASE and H. NISHIMURA (2006), "Autonomous Traveling Control of the 'TAO Aicle' Intelligent Wheelchair", *Proceed IEEE/RSJ Int Conf IROS*, pp. 4322-4327.
MUSHA, T., T. ASADA, F. YAMASHITA, T. KINOSHITA, Z. CHEN, H. MATSUDA *et al.* (2002), "A New EEG Method for Estimating Cortical Neuronal Impairment That Is Sensitive to Early-Stage Alzheimer's Disease", *Neurophysiol Clin* 113, pp. 1052-1058.
NeCoRo. http://wwwnecorocom/homehtml, in press.
NISHIKAZE, O. *et al.* (1995) "Distortion of Adaptation (Wear & Tear and Repair & Recovery)-Urine 17- KS-Sulfates and Psychosocial Atressin", *Humans-Job Stress Res* 3, pp. 55-64.
PaPeRo. http://wwwincxneccojp/robot/robotcenterhtml.
PINES (1981), "The Burnout Measure", paper presented at the National Conference on Burnout in the Human Services, Philadelphia, November.
RICHESON, N. E. (2003), "Effects of Animal-Assisted Therapy on Agitated Behaviors and Social interactions of Older Adults with Dementia", *Am J Alzheim. Dis. Other Dement* 18(6), pp. 353-358.
Robo-One. http://wwwrobo-onecom.
SAITO, T., T. SHIBATA, K. WADA and K. TANIE (2002), "Examination of Change of Stress Reaction by Urinary Tests of Elderly before and after Introduction of Mental Commit Robot to an Elderly Institution", *Proceed. Int. Symp. AROB* 1, pp. 316-319.
SELYE, H. (1970), "Stress and Aging", *Journal of American Geriatric Society* 18, pp. 669-676.
SHIBATA, T. (2004), "An Overview of Human Interactive Robots for Psychological Enrichment", *Proceed. IEEE* 92(11), pp. 1749-1758.
SHIBATA, T. (2004), "Ubiquitous Surface Tactile Sensor",*TExCRA*, p. 56.
SHIBATA, T., K. INOUE and R. IRIE (1996), "Emotional Robot for Intelligent System - Artificial Emotional Creature Project", *Proceed. IEEE RO-MAN,* pp. 466-471.
SHIBATA, T. and R. IRIE (1997), "Artificial Emotional Creature for Human-Robot Interaction - A New Direction for Intelligent System", *Proceed IEEE/ASME Int Conf AIM* 47.
SHIBATA, T., T. MITSUI, K. WADA and K. TANIE (2002), "Subjective Evaluation of Seal Robot: Paro - Tabulation and Analysis of Questionnaire Results", *J Robot Mechatron* 14(1), pp. 13-19.
SHIBATA, T., T. MITSUI, K. WADA, A. TOUDA, T. KUMASAKA, K. TAGAMI *et al.* (2001), "Mental Commit Robot and its Application to Therapy of Children", *Proceed. IEEE/ASME Int. Conf. AIM*, pp. 1053-1058.
SHIBATA, T. and K. TANIE (2000), "Influence of A-Priori Knowledge in Subjective Interpretation and Evaluation by Short-Term Interaction with Mental Commit Robot", *Proceed.IEEE/RSJ Int Conf IROS* 1, pp. 169-174.
SHIBATA, T., T. TASHIMA and K. TANIE (1999), "Subjective Interpretation of Emotional Behavior through Physical Interaction between Human and Robot", *Proceed. IEEE Int. Conf. SMC*, pp. 1024-1029.
SHIBATA, T., K. WADA, Y. IKEDA and S. SABANOVIC (2009), "Cross-Cultural Studies on Subjective Evaluation of Seal Robot", *Advanced Robotics,* in press.
SHIBATA, T., K. WADA, T. SAITO and K. TANIE (2001), "Robot Assisted Activity for Senior People at Day Service Center", *Proceed. Int. Conf. ITM,* pp. 71-76.
TAMURA, T., S. YONEMITSU, A. ITOH, D. OIKAWA, A. KAWAKAMI, Y. HIGASHI *et al.* (2004), "Is an Entertainment Robot Useful in the Care of Elderly People With Severe Dementia?", *J Gerontol. A Biol Sci. Med. Sci. 2004* 59A(1), pp. 83-85.
WADA, K. and T. SHIBATA (2007), "Living with Seal Robots – Its Socio-Psychological and Physiological Influences on the Elderly in a Care House", *IEEE Trans. Robot* 23(5), pp. 972-980.
WADA, K., T. SHIBATA, T. MUSHA and S. KIMURA (2008), "Robot Therapy for Elders Affected by Dementia", *IEEE Eng Med Biol Mag* 27(4), pp. 53-60.
WADA, K., T. SHIBATA, T. SAITO, K. SAKAMOTO and K. TANIE (2005), "Psychological and Social Effects of One Year Robot-Assisted Activity on Elderly People at a Health Service Facility for the Aged", *Proc. IEEE ICRA,* pp. 2796-2801.
WADA, K., T. SHIBATA, T. SAITO and K. TANIE (2004), "Effects of Robot-Assisted Activity for Elderly People and Nurses at a Day Service Center", *Proceed. IEEE* 92(11), pp. 1780-1788.
WADA, K., T. SHIBATA, T. SAITO and K. TANIE (2004), "Psychological and Social Effects in Long-Term Experiment of Robot-Assisted Activity to Elderly People at a Health Service Facility for the Aged", *Proceed. IEEE/RSJ Int. Conf. IROS,* pp. 3068-3073.
WPA/PTD Educational Program on Depressive Disorders "Depressive Disorders in Older Persons", Module 3: Depressive Disorders in Older Persons http://www.wpanet.org/education/ed-program-guidelines.html
YESAVAGE, J. A. (1988), "Geriatric Depression Scale", *J Psychopharmacology Bulletin* 24(4), pp. 709-711.

BBVA

Innovation: Changing the Face of Disability

Hugh Herr and Ernesto Martinez-Villalpando
MIT Media Lab

Approximately 650 million people in the world suffer from some type of disability and as the population ages this figure is expected to increase. Those afflicted with a physical, emotional or cognitive disability face a myriad of serious and debilitating challenges. Fortunately, the modern explosion in scientific and technological innovations provides an extraordinary opportunity to deliver profound improvements to their quality of life. Moreover, the same cutting-edge technology that can minimize or eliminate the adverse effects of disabling conditions can also be used to expand human abilities and transcend the ordinary limits of the human condition.

At the Massachusetts Institute of Technology's (MIT) Media Laboratory, the Biomechatronics Group and the Affective Computing Group are focusing on developing novel technologies that can deeply impact people's lives at the physical and emotional levels. The Biomechatronics Group seeks to understand the basic principles of biological locomotion in order to develop both rehabilitation technologies that restore functionality to the physically challenged and augmentation technologies that amplify the physical capabilities of healthy able-bodied individuals. The Affective Computing group works towards the development of technologies to expand our understanding of affect and its role in the human experience, with a focus on autism research and therapy. The interdisciplinary nature of the work of both groups integrates a broad gamut of disciplines, ranging from medicine to engineering.

The Biomechatronics Group and Affective Computing Group are part of MIT's Human 2.0 initiative to mitigate the effects of disability and redefine the limits of human capability.

BIOMECHATRONICS: DISABLED PEOPLE OR DISABLED TECHNOLOGIES?

Though often taken for granted, walking remains essential in modern life, as stairs, uneven terrain, and other obstacles easily conquered by legs but not wheels remain ubiquitous. The loss or disability of a leg tremendously impacts quality of life and patients strive to regain or retain the ability to walk even in the presence of severe impairment.

In the United States there are more than 26 million people with physical disabilities including more than 1.7 million (more than 1 in 200) living with limb loss (NLLIC, 2008). In order to restore lost limb functions, prosthetic and orthotic technology is generally required. The need for rehabilitation and prosthetic technology is latent, as the total number of persons with an amputation and using a prosthesis is expected to reach 2.4 million by the year 2020 (Ziegler-Graham, 2008).

Currently commercially-available technologies for lower limb amputees are still far from providing fully functional replacements of biological legs. Even with the most advanced prosthetic systems available on the market, amputees still exhibit clinical problems associated with lack of adequate mobility. These include gait asymmetry, instability, decreased walking speeds and higher energy requirements. Together these gait pathologies result in significant pain and walking fatigue for lower limb amputees (Postema *et al.,* 1997).

Although the pain felt at the residual limb corresponds to the behavior of the entire prosthetic system (i.e. from the liner and socket interface to the pylon and the rest of the prosthetic components), it is particularly associated with the coupling between the residual limb and the prosthetic leg. The imperfect coupling allows relative motion between the socket and the femur stump caused by the compression of soft tissue. This motion is uncomfortable for the amputee and causes a lack of confidence to apply large forces to the prosthetic leg. In addition, the relatively short moment arm between the hip joint and the socket reduces the force that the hip muscles can apply to the artificial limb (Whittle, 1991).

Recent advances in socket technology have reduced pain in patients by focusing on cushioning, a primary contributor to comfort. Such technologies cover a large spectrum, from gel liners and vacuum-assisted sockets to modern interfaces that rely on residual limb laser scanning and computer-aided manufacturing. Two particular technologies that have proved to be successful in pain reduction have been shock absorbing pylons and dynamic elastic response (DER) prosthetic feet (Perry *et al.,* 1992). The damping and compliance features they provide have made them popular in most of the commercially available prosthetic systems. Despite their success in preference among amputees, abnormal gait patterns and associated with walking fatigue are still prevalent.

Walking fatigue is synonymous with higher metabolic expenditure and is a common affliction of lower-limb amputees. Walking fatigue in lower-limb amputees is considerably higher than in their matched able-bodied counterparts at comparable speeds. Measures of metabolic expenditure during walking are commonly obtained by analyzing oxygen level consumptions. For unilateral below-the-knee amputees, the rate of oxygen consumption is 20-30% higher (Herbert *et al.,* 1994; Molen, 1973) than that for healthy persons with no impairments, and for above-knee amputees this rate increases by an additional 25% (James, 1973; Waters and Mulroy, 1999).

Conventional lower-limb prostheses, despite their damping and compliance features, have not provided a real metabolic advantage for amputees (Lehmann *et al.,* 1993; Torburn *et al.,* 1990; Colborne *et al.,* 1992; Huang *et al.,* 2000; Thomas *et al.,* 2000). In addition to higher energetic requirements, lower-limb amputees show a reduction in their self-selected speed, and in consequence they present overall diminished endurance.

> “Currently commercially-available technologies for lower limb amputees are still far from providing fully functional replacements of biological legs”

Prosthetic systems ideally need to fulfill a diverse set of requirements in order to restore the biological behavior of normal and healthy limbs. For the Biomechatronics Group, the biomechanics of normal walking provide a basis for the design and development of new actuated artificial limbs. This unique biomimetic approach to the design and development of these prostheses shows promise in improving amputees' gait symmetry, walking speed and metabolic requirements while enhancing the adaptation to the particular amputee's gait.

One of the objectives of the Biomechatronics Group is to develop sophisticated modular biomimetic leg prosthesis for lower-limb amputees that is capable of restoring the functionality of the ankle and knee joints of the intact human leg and fully emulating their natural behavior. This task poses many challenges for researchers as they investigate novel electromechanical designs and control strategies that can adequately integrate and adapt to the patients' needs. The complete robotic lower l3 limb is comprised of two modular robotic joint prostheses: a powered ankle-foot and robotic knee prosthesis.

Robotic Ankle-Foot Prosthesis

The human ankle joint is essential to locomotion because it provides a significant amount of energy to push the body off the ground and propel it forward during walking, especially at moderate to fast speeds (Winter, 1983; Palmer, 2002; Gates, 2004). For transtibial (below-the-knee) amputees, the loss of this energy generation at the ankle produces an abnormal asymmetric gait, with higher metabolic energy requirements and slower speeds. Additionally, the mechanical behavior of commercially available ankle-foot prostheses greatly differs from that of a healthy human ankle-foot. Even though most of these prostheses offer some compliance and function as initial and terminal rockers due to their shape, they cannot provide the amount of external energy required in walking, making them inadequate in replicating the natural ankle's flexibility and actuation (Whittle, 1991).

In order to overcome the disadvantages of current prosthetic technologies for below-knee amputees, the Biomechatronics Group has developed the world's first robotic ankle-foot prosthesis that can successfully recreate the actions of the biological lower leg (Au and Herr, 2006; Au *et al.,* 2007). Using advanced biologically-inspired design and intelligent computer algorithms, this novel device can propel an amputee forward while easily adapting to changes in ambulation speed and the walking environment. This artificial ankle-foot prosthesis allows amputees to enjoy a natural human gait over level ground, stairs, ramps, and even uneven terrain. Moreover, the device's low weight and biological form-factor make it comfortable to wear and inconspicuous to even the trained eye. Most importantly, this innovative device reduces the rate of oxygen consumption in walking amputees by up to 20% relative to conventional prosthetic devices (Au *et al.,* 2009).

The success of the active ankle prosthesis derives from the Biomechatronics Group's commitment to biomimetic design. The mechanical design of this motorized device imitates the biological structures of the ankle joint by using elastic elements and flexible materials in similar roles to those of the tendons and ligaments of the human joint. This exploitation of elastic elements reduces the weight of the necessary motor and minimizes the overall power that this battery-operated system requires, allowing an amputee to walk all day on a single charge.

This cutting-edge bionic research device has been so successful that it was awarded *Time Magazine*'s Best Invention of the year in 2007 and led to the creation of iWalk LLC., a start-up company commercializing this prototype.

Robotic Knee Prosthesis

For above-knee amputees, a particular source of pathological gait while wearing conventional prostheses is the lack of accurate control of the knee joint, particularly while the leg is swinging during each step. The knee cannot be allowed to swing freely because it will extend too rapidly and stop suddenly as it reaches full extension. On the other hand, the knee joint cannot be so rigid that it does not bend in response to dynamics; such rigidity would result in a large increase in the amount of energy required by the patient to go from one step to the next. To prevent these extreme cases, several prosthetic knees that behave as dampers (e.g. energy-dissipation mechanisms) have been developed using friction, hydraulic, pneumatic, or electro-mechanical systems. Some have been designed as variable damping devices which adapt to angle, speed and direction of motion. These mechanisms have partially addressed abnormal gait patterns in amputees (Whittle, 1991), but have not yet been able to mimic fully the complex behavior of the knee joint.

Building on the work that led to the world's first powered ankle, the Biomechatronics group has continued its line of innovation in bionic limbs by developing a state-of-the-art robotic knee

Figure 1. Biomechatronics' Robotic Ankle-Foot Prosthesis (*photograph by Webb Chappell ©MIT Media Lab*) and most recent prototype by iWalk, LLC.

joint prosthesis which overcomes the limitations of conventional prosthetic knees. This prosthesis is capable of replicating the behavior of the biological knee joint while seamlessly interfacing with the powered ankle, producing a full artificial lower-limb prosthesis (Martinez-Villalpando *et al.,* 2008; Martinez-Villalpando and Herr, 2010).

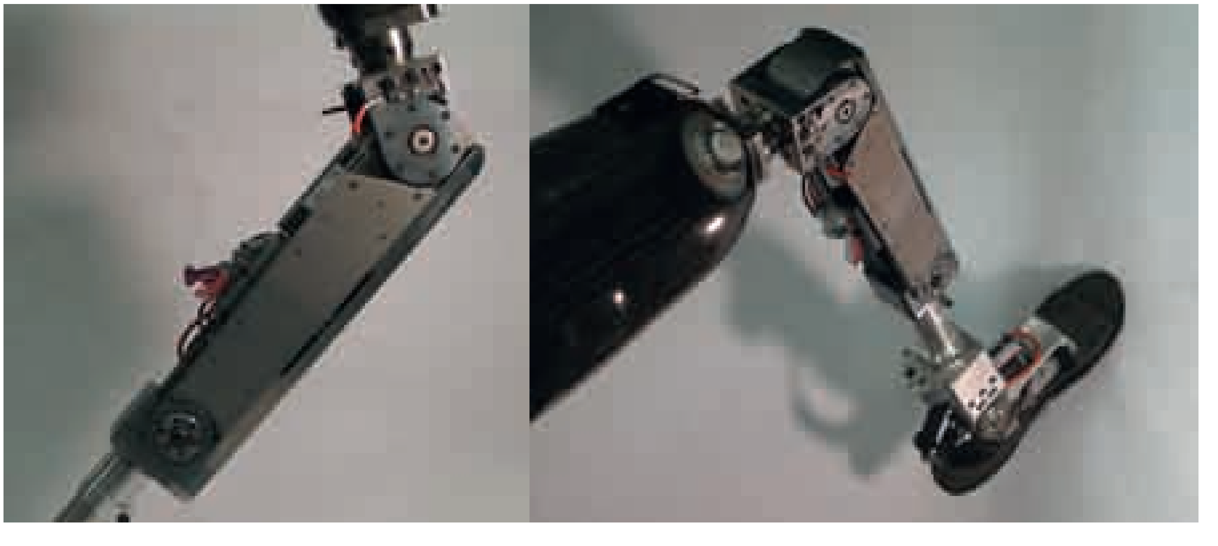
The active knee prosthesis is a novel motorized device with a unique biomimetic electro-mechanical design. The artificial knee mimics the functionality of the musculo-skeletal structures around an intact biological knee joint, producing a system that, like the artificial ankle, is small, light-weight, and efficient. Its design incorporates a microcomputer and a sophisticated sensory suite that enables an artificial intelligence capable of inferring the intentions of the amputee. The advanced design and control of this prosthesis aims to improve amputee gait beyond what other commercially available prostheses can offer, not only while walking on even ground but also while traversing difficult terrain, including ramps and stairs. The integration of the robotic knee and ankle prosthesis into a single prosthetic system yields the world's most advanced powered artificial lower limb for transfemoral amputees. Together, the active knee and ankle are expected to have great clinical impact while their novel architectures contribute to the development of future integral assistive technologies that adapt to the needs of the disabled.

Exoskeletons

Physical disabilities that often result in leg weakness include lower-extremity amputation, spinal-cord impairment (SCI), multiple sclerosis (MS) and stroke. For individuals who have suffered partial leg paralysis resulting from neurological pathology, the use of exoskeleton technology will offer a dramatic improvement in mobility capabilities over conventional leg orthotic technology.

The Biomechatronics group is leveraging its understanding of human locomotion and its experience in prosthetic limb design towards the development of exoskeletons (Walsh *et al.,* 2006; 2007). These wearable assistive structures help augment human mobility, increase human endurance and

Figure 2. The Biomechatronics Groups' Robotic Knee Prosthesis



assist physically-challenged persons. The group's walking exoskeleton is an orthotic system that works in parallel to the body, transmitting forces between the ground and the user's torso during standing and walking, effectively reducing the portion of body weight borne by the legs and making it easier for a disabled person to stand and to walk. Because the exoskeleton offers support normally provided by biological legs, physically disabled people suffering from leg weakness may walk with confidence while wearing it (Dollar and Herr, 2007, 2008).

In particular, the exoskeleton work developed at Biomechatronics Group looks into the passive dynamics of human walking in order to create lighter and more efficient devices with three specific goals in mind. Firstly, the exoskeleton currently under development aims to be the first wearable system that demonstrates a reduction in human energy usage during walking. Secondly, the exoskeleton should serve in potentially life-saving occupations, increasing the user's endurance while reducing damaging loads on the knee and ankle. These potential users include active soldiers and firefighters, among others, who perform activities that require brisk movement over varying terrain while burdened with significant loads. Finally, this technology aims to assist impaired human mobility. This exoskeleton architecture could be modified into a walking orthosis which permits an active lifestyle by reducing load on injured joints while providing the necessary support for normal walking to patients with otherwise limited mobility.

Figure 3. The Biomechatronics Group's load-carrying exoskeleton

AFFECTIVE COMPUTING: THE AUTISM CHALLENGE

Autism Spectrum Disorders (ASD) are a collection of neuro-developmental disorders characterized by qualitative impairments in socialization, communication, and circumscribed interests, including stereotypical behavior patterns and behavioral rigidity to changes in routines (APA, 1994). Current studies of ASD suggest a rate as high as 1 in 110 in children by the age of 8 years in the United States (CDCP, 2009). ASDs typically manifest in infancy and persist throughout the lifespan. These disorders have a profound impact on families and often result in enormous emotional and financial costs. For instance, recent estimates suggest that the societal costs in the United States to care for all individuals diagnosed each year over their lifetime approaches $35 billion (Gantz, 2007). ASDs clearly represent an emerging public health problem (Newschaffer *et al.,* 2003).

Through the Affective Computing group and the Autism and Communication

Technology Initiative at the MIT Media Laboratory, a variety of innovative technologies are being developed to better understand and support individuals with ASD in natural environments. Three of these applications, briefly reviewed in the following paragraphs, include: 1. automatically detecting stereotypical motor movements using wireless accelerometers and pattern recognition algorithms; 2. developing unobtrusive, wireless measures of physiological arousal; and 3. creating a suite of wearable, wireless technologies that enable the capture, real-time analysis, and sharing of *in situ* social-emotional cues from faces, voices, and gestures of self and/or interaction partners.

Sensor-Enabled Detection of Stereotypical Motor Movements

Stereotypical motor movements (SMMs) are generally defined as repetitive motor sequences that appear to an observer to be invariant in form and without any obvious eliciting stimulus or adaptive function. Several SMMs have been identified, the most prevalent among them being body rocking, mouthing, and complex hand and finger movements (Lewis and Bodfish, 1998). SMMs occur frequently in people with mental and developmental disabilities, genetic syndromes (Bodfish *et al.,* 2000), and less frequently in normally developing children and adults.

While investigations of ASD have increased in recent years in response to growing awareness of the high prevalence rates, the majority of this work focuses on social and communication deficits, rather than on restricted and repetitive behavior. This is a potential problem given the high prevalence of SMMs reported in individuals with ASD. Also, when severe, SMMs can present several difficulties for individuals with ASD and their families. First, persons with ASD often engage in SMMs. Preventing or stopping these movements can be problematic as individuals with ASD may become anxious, agitated, or aggressive if they are interrupted (Gordon, 2000). Second, if unregulated, SMMs can become the dominant behavior in an individual with ASD's repertoire and interfere with the acquisition of new skills and the performance of established skills (Koegel and Covert, 1972). Third, engagement in these movements is socially inappropriate and stigmatizing and can complicate social integration in school and community settings (Jones *et al.,* 1990). Finally, SMMs are thought to lead to self-injurious behavior under certain environmental conditions (Kennedy, 2002).

To better measure, understand, and remediate this complex class of behavior, we are developing an innovative system for automatically recognizing and monitoring SMM. Our system uses a miniature sensory suite that is comfortably worn on an individual's wrists and torso and transmits motion data to a mobile phone. Pattern-recognition algorithms running on the phone receive these motion data streams wirelessly, compute a variety of characteristic features, and automatically detect SMM topography, onset, offset, frequency, duration, and intensity (Munguia-Tapia *et al.,* 2004). Currently, this system has been able to correctly identify stereotypical body rocking, hand flapping, and head hitting approximately 90% of the time across six individuals with ASD in both laboratory and classroom settings (Albinali *et al.,* 2009).

There are several potential benefits associated with this novel system. Easily automating SMM detection could free a human observer to concentrate on and

Figure 4. The MIT's 3-axis wireless-accelerometer sensors housed in plastic cases with external battery holder. The cases can be worn on the wrists using elastic armbands.

note environmental antecedents and consequences necessary to determine what functional relations exist for this perplexing and often disruptive class of behavior. The system could also be used as an outcome measure to facilitate efficacy studies of behavioral and pharmacological interventions intended to decrease the incidence or severity of SMM. Finally, with minor modifications, the system could be programmed to serve as an intervention tool by providing real-time feedback to individuals with ASD and/or their caregivers when SMMs are detected.

Unobtrusive, Wireless Measures of Physiological Arousal

The Autonomic Nervous System (ANS) is a control system in the body with far-reaching influences, including maintenance of heart rate, digestion, respiration rate, and perspiration that mediates regulation of emotion, shifting of attention, sleep, signaling of anticipation and salience, biasing of memory, and more.

A number of investigators over the past thirty years have recorded ANS activity in individuals with ASD to assess physiological responsivity during attention and habituation tasks, while exposed to social and sensory stimuli, and when engaged in self-injurious and repetitive behaviors. Unfortunately, however, there are several methodological issues associated with these studies that cast doubt on the reliability, validity, and generalizability of the data obtained. For instance, the majority of ANS studies to date use obtrusive equipment that requires individuals to sit still while multiple wires are adhered to their chest or fingers, limiting the number of participants who can comply with the procedures and thus contribute data to a study. ANS observations are also undertaken primarily in unfamiliar research laboratories that are potentially stress-inducing, and are often limited to short intervals of measurement that may or may not represent a person's true ANS patterns when going about everyday activities. Data from these studies are also often averaged across persons so that no individual profiles are retained, obscuring the heterogeneity of response patterns across individuals.

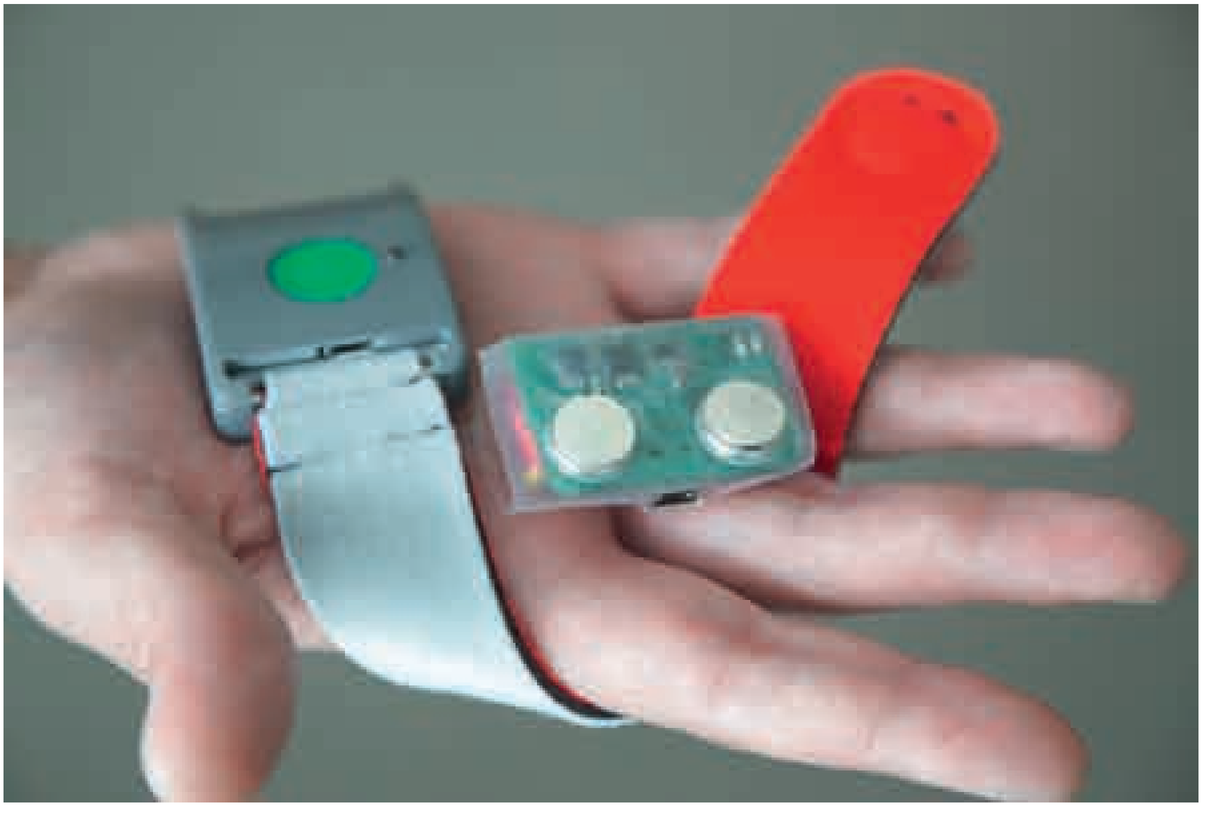
To overcome some of these methodological problems, a novel technology platform is being developed for sensing sympathetic and parasympathetic autonomic data comfortably off the wrist and ankle without wires or boxes (Fletcher *et al.,* 2010; Poh *et al.,* 2010). The system captures: 1. electrical conductivity of the skin, which provides a sensitive measure of changes in sympathetic arousal associated with emotion, cognition and attention; 2. heart rate and heart rate variability that provides information related to the sympathetic and parasympathetic branches of the ANS; 3. temperature; and 4. motor movement and posture changes through 3-axis accelerometry. The 3-axis accelerometer and temperature sensors provide information about a person's activity and account for the influence of motion and environmental temperature on electrical conductivity of the skin and cardiovascular signals.

Monitoring autonomic reactivity using comfortable, wireless, wearable packages could enable new *in situ* experimental paradigms and address some of the shortcomings associated with traditional methods of recording the ANS in persons with ASD. For instance, these sensors could enable longitudinal studies of individuals that yield data beyond the traditional "snapshot" timeframe, providing new insights on within-person, within-group, and across-group differences over time, and capturing phenomena of interest that are hard to replicate in laboratory settings, e.g., panic attacks. Measuring and communicating ANS patterns that precede, co-occur, and follow an event could also provide rich data enabling new ways to anticipate, respond to, and ultimately prevent problem behaviors (e.g., self-injury, aggression towards others).

Interactive Social-Emotional Toolkit (iSET)

Many first-hand accounts from people with ASD highlight the challenges of interacting

Figure 5. MIT Media Lab wearable EDA sensor. (Right) Sensor with disposable Ag/AgCl electrodes attached. (Left) Elastic strap with hardcase form factor that contains the sensor.



socially and difficulties inherent in the real-time processing of high-speed, complex, and unpredictable information like nonverbal cues (e.g., facial expressions) or making eye contact while processing language at the same time. Difficulties such as the following are also well-documented in a large body of empirical literature: 1. difficulty relating to other peoples' nonverbal cues and mental states (Baron-Cohen, 1995); 2. atypical eye-gaze processing (Klin *et al.,* 2002); 3. difficulty understanding and expressing one's own feelings (Hill *et al.,* 2004); and 4. trouble gauging the interests of others in conversation (Klin *et al.,* 2000). These challenges affect interaction partners as well, making it hard for family members and others to understand what the person is trying to communicate.

> "Utilizing recent advances in pervasive and ubiquitous computing, sensors, and camera technology, it is now possible to have a range of on-body sensors that communicate to a portable device such as a mobile phone or ultra mobile PC"

Utilizing recent advances in pervasive and ubiquitous computing, sensors, and camera technology, it is now possible to have a range of on-body sensors that communicate to a portable device such as a mobile phone or ultra mobile PC. Building on these advances, we are developing an interactive social-emotional toolkit (iSET) (figure 4): a suite of wearable, wireless technologies that enable the capture, real-time analysis, and sharing of *in situ* social-emotional cues from faces, voices, and gestures of self and/or interaction partner(s). The technology components of iSET include a wearable camera that can be worn facing the wearer (Self-Cam) or facing outward (Head-Cam). The captured video is processed using real-time video-pattern analysis algorithms and is tagged at multiple granularities (facial actions, communicative facial/head gestures, and emotions) (el Kaliouby and Robinson, 2005).

The iSET project makes these wearable components available and accessible to individuals on the autism spectrum in the hope that it will allow this population to systemize, quantify, and reflect on their social interactions, which otherwise may seem confusing, overwhelming, and beyond their control. iSET is also designed to be fun, turning social interactions into a stimulating game that might motivate participants to engage in communication. The data and analyses offered by iSET also facilitate the sharing of social experiences with family members, teachers, and friends, and thus are inherently social.

Currently this system is being iteratively tested in the following scenarios at a large school for individuals with ASD:

Face and eye contact. "Head-Cam" or "Third-Eye" is a wearable camera that points outward and is aligned with the wearer's field of vision (Lee *et al.,* 2008). The video stream is input to face-detection software that quantifies how much face-contact happens in a natural conversation.

Systemizing social-emotional cues in self and others. Many individuals on the autism

Figure 6. *A* student on the autism spectrum using iSET with his teacher to capture, tag, and analyze facial expressions.

spectrum report difficulties expressing themselves in socially appropriate ways, and find it hard to identify their own feelings, as well as the feelings of others. "Self-Cam" is designed to help a person re-experience and reflect on how he/she appears to others (Teeters, 2007).

Learning what matters. While many interventions address the problem of recognition of social-emotional cues, very few teach individuals on the autism spectrum how to identify the cues to which it is most important to pay attention. Without this aspect of social processing, a person might attempt to process every single instance of social cues, an undoubtedly time-consuming and cognitive overloading process that detracts from that person's ability to respond in real time to his/her interaction partner. "Expressions Hunt" is a situated game we are developing in which individuals are given the task of eliciting and capturing various facial expressions such as smiles or frowns from others using a wearable camera. In this game, wearers have to think about eliciting, not just recognizing and capturing a state.

REDEFINING THE HUMAN CONDITION

We live in exciting times, as unprecedented progress in science and technology redefines human disability. Institutional collaboration and the integration of a broad range of disciplines are producing sophisticated solutions which allow people with once-debilitating physical and mental health conditions to lead healthy full lives. Meanwhile, novel technologies that create intimate connections between man and machine are augmenting human abilities beyond natural limits. Without a doubt, the contributions of leading scientists and engineers, including those at MIT's

Biomechatronics and Affective Computing Groups, are forcing society to reexamine disability and what it means to be human.

ACKNOWLEDGEMENTS

The authors would like to express their utmost gratitude and appreciation to Dr. Mattew S. Goodwin (director of clinical research at the MIT Media Lab and associate director of research at the Groden Center, an institute for autism spectrum disorders in Providence, RI) and to Professor Rosalind W. Picard, Sc.D**.** (founder and director of the Affective Computing Research Group) for their invaluable contributions to the content of this article.

REFERENCES

http://affect.media.mit.edu/
http://biomech.media.mit.edu/
http://www.iwalk.com/
http://www.disabled-world.com
http://www.media.mit.edu/research/autism-communication-technology-initiative

ALBINALI, F, M. S. GOODWIN and S. S. INTILLE (2009), "Recognizing Stereotypical Motor Movements in the Laboratory and Classroom: A Case Study with Children on the Autism Spectrum", *Proceedings of the 11th International Conference on Ubiquitous Computing*, New York: ACM Press, p. 80.

American Psychiatric Association (1994), *Diagnostic and statistical manual of mental disorders,* 4th ed., Washington, DC: Author.

AU, S., and H. HERR (2006), "Initial Experimental Study on Dynamic Interaction between an Amputee and a Powered Ankle-foot Prosthesis", *Workshop on Dynamic Walking: Mechanics and Control of Human and Robot Locomotion*, Ann Arbor, MI.

AU, S., J. WEBER, E. MARTINEZ-VILLAPANDO, and H. HERR (2007), "Powered Ankle-Foot Prosthesis for the Improvement of Amputee Ambulation", *IEEE Engineering in Medicine and Biology International Conference,* August 23-26, Lyon, France, pp. 3020-3026.

AU, S., M. BERNIKER and H. HERR, (2008), "Powered Ankle-Foot Prosthesis to Assist Level-Ground and Stair-Descent Gaits", *Neural Networks* 21, pp. 654-666.

AU, S., J. WEBER and H. HERR (2009) "Powered Ankle-foot Prosthesis Improves Walking Metabolic Economy", *IEEE Transactions on Robotics* 25, pp. 51-66.

BARON-COHEN, S. (1995), *Mindblindness,* Cambridge, MA: The MIT Press.

BODFISH, J. W., F. J. SYMONS, D. E. PARKER, and M. H. LEWIS (2000), "Varieties of Repetitive Behaviors in Autism: Comparisons to Mental Retardation", *Journal of Autism and Developmental Disorders* 30, pp. 237-243.

Centers for Disease Control and Prevention (2009), *Prevalence of Autism Spectrum Disorders—Autism and Developmental Disabilities Monitoring Network, United States,* Surveillance Summaries (December 18, 2009), *MMWR* 58 (no. SS-10).

COLBORNE, G. R., S. NAUMANN, P. E. LONGMUIR and D. BERBRAYER (1992), "Analysis of Mechanical and Metabolic Factors in the Gait of Congenital Below-Knee Amputees: a Comparison of the SACH and Seattle Feet", *American Journal of Physical Medicine and Rehabilitation* 71, pp. 272-278.

DOLLAR, A., H. HERR (2007), "Active Orthoses for the Lower Limbs: Challenges and State-of-the-Art ", *Proceedings of the 2007 IEEE International Conference on Rehabilitation Robotics (ICORR)*, Noordwijk, Netherlands.

DOLLAR, A., H. HERR (2008), "Lower Extremity Exoskeletons and Active Orthoses: Challenges and State-of- the Art", *IEEE Transactions on Robotics* 24(1).

EILENBERG, M. F., H. GEYER and H. HERR (2010), "Control of a Powered Ankle-Foot Prosthesis Based on a Neuromuscular Model", *Transactions on Neural Systems and Rehabilitation Engineering*.

EL KALIOUBY, R. and P. ROBINSON (2005), "Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures", *Real-Time Vision for Human-Computer Interaction*, pp. 181-200, Springer-Verlag.

FLETCHER, R. R., K. DOBSON, M. S. GOODWIN, H. EYDGAHI, O. WILDER-SMITH, D. FERNHOLZ, and R. W. PICARD (2010), "iCalm: Wearable Sensor and Network Architecture for Wirelessly Communicating and Logging Autonomic Activity", *IEEE Transactions on Information Technology in Biomedicine* 14, pp. 215-223.

GANTZ, M. L. (2007),"The Lifetime Distribution of the Incremental Societal Costs of Autism", *Archives of Pediatrics and Adolescent Medicine* 161, pp. 343-349.

GATES, D. H. (2004), "Characterizing the Ankle Function during Stair ascent, Descent and Level Walking for Ankle Prosthesis and Orthosis Design", Boston, MA: Boston University, Master's Thesis.

GORDON, C. T. (2000), Commentary: "Considerations on the Pharmacological Treatment of Compulsions and Stereotypes with Serotonin Reuptake Inhibitors in Pervasive Developmental Disorders", *Journal of Autism and Developmental Disorders* 30, pp. 437-438

HERBERT, L. M., J. R. ENGSBERG, K. G. TEDFORD and S. K. GRIMSTON (1994), "A Comparison of Oxygen Consumption during Walking between Children with and without Below-Knee Amputations", *Physical Therapy* 74: pp. 943-950.

HILL, E. L., S. BERTHOZ and U. FRITH (2004), Brief report: "Cognitive Processing of own Emotions in Individuals with Autistic Spectrum Disorder and in their Relatives", *Journal of Autism and Developmental Disorders* 34, pp. 229-235.

HUANG, G. F., Y. L. CHOU and F. C. SU (2000), "Gait Analysis and Energy Consumption of Below-Knee Amputees Wearing Three Different Prosthetic Feet", *Gait Posture* 12, pp. 162-168.

JAMES, U. (1973), "Oxygen Uptake and Heart Rate During Prosthetic Walking in Healthy Male Unilateral Above-Knee Amputees", *Scandinavian Journal of Rehabilitation Medicine* 5: pp. 71-80.

JONES, R. S. P., D. WINT and N. C. ELLIS (1990), "The Social Effects of Stereotyped Behavior", *Journal of Mental Deficiency Research* 34, pp. 261-268.

KENNEDY, C. H. (2002), "Evolution of Stereotypy into Self-Injury, in S. R. Schroeder, M. L., Oster-Granite *et al.,* (eds.), *Self-Injurious Behavior: Gene-Brain Behavior Relationships*, Washington, D.C.: American Psychological Association, pp. 133-143.

KLIN, A., W. JONES, R. SCHULTZ, F. VOLKMAR and D. COHEN (2002), "Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals with Autism", *Archives of General Psychiatry* 59, pp. 809-816.

KLIN, A., F. VOLKMAR and S. SPARROW (2000), *Asperger Syndrome,* New York: Guilford Press.

KOEGEL, R. L., and A. COVERT (1972), "The Relationship of Self-Stimulation to Learning in Autistic Children", *Journal of Applied Behavior Analysis* 5, pp. 381-387.

LEE, C. H., R. MORRIS, M. GOODWIN and R. W. PICARD (2008), "Lessons Learned from a Pilot Study Quantifying Face Contact and Skin Conductance in Teens with Asperger Syndrome", Extended Abstract of *CHI 2008*.

LEHMANN, J. F., R. PRICE, S. BOSWELL-BESSETTE, A. DRALLE and K. QUESTAD (1993), "Comprehensive Analysis of Dynamic Elastic Response Feet: Seattle Ankle/Lite Foot versus SACH Foot", *Archives Physical Medicine and Rehabilitation* 74: pp. 853-861.

LEWIS, M. H., and J. W. BODFISH (1998), "Repetitive Behavior Disorders in Autism", *Mental Retardation and Developmental Disabilities Research Reviews* 4, pp. 80-89.

MARTINEZ-VILLALPANDO, E. C., and H. HERR (2009), "Agonist-Antagonist Active Knee Prosthesis: a Preliminary Study in Level-Ground Walking", *Journal of Rehabilitation, Research and Development* 46(3), pp. 361-373.

MARTINEZ-VILLALPANDO, E. C., J. WEBER, G. ELLIOTT and H. HERR (2008), "Design of an Agonist- Antagonist Active Knee Prosthesis", *IEEE BIORobotics Conf.*, Scottsdale, AZ.

MARTINEZ-VILLALPANDO, E. C., J. WEBER, G. ELLIOTT and H. HERR (2008), "Biomimetic Prosthetic Knee using Antagonistic Muscle-Like Activation ", *ASME International Mechanical Engineering Congress and Exposition (IMECE)*, Boston, MA.

MOLEN, N. H. (1973), "Energy-Speed Relation of Below-Knee Amputees Walking on a Motor-Driven Treadmill", *International Z Angew Physiology* 31, pp.173-185.

MUNGUIA TAPIA, E., S. S. INTILLE and K. LARSON (2004), "Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors", A. Ferscha and F. Mattern (eds.), *Proceedings of PERVASIVE 2004*, vol. LNCS 3001, pp. 158-175, Berlin Heidelberg: Springer-Verlag.

NATIONAL LIMB LOSS INFORMATION CENTER (2008), "Amputation Statistics by Cause: Limb Loss in the United States", *Amputee Coalition of America,* Knoxville, TN.

NEWSCHAFFER, C. J., and L. K. CURRAN (2003), "Autism: An Emerging Public Health Problem", *Public Health Reports* 118, pp. 393-399.

PALMER, M. (2002), "Sagittal Plane Characterization of Normal Human Ankle Function Across a Range of Walking Gait Speeds", Boston, MA: Massachusetts Institute of Technology, Master's Thesis.

PERRY, J. (1992), "Gait Analysis Normal and Pathological Function", *Slack, Inc*., Thorofare, NJ.

POH, M., N. C. SWENSON and R. W. PICARD (2010), "A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity", *IEEE Transactions on Biomedical Engineering* 57, pp. 1243-1252.

POSTEMA, K., H. J. HERMENS, J. DE VRIES, H. F. KOOPMAN and W. H. EISMA (1997), "Energy Storage and Release of Prosthetic Feet, Part 1: Biomechanical Analysis Related to User Benefits", *Journal Prosthetics and Orthotics International* 21, pp.17-27.

TEETERS, A. (2007), "Use of a Wearable Camera System in Conversation: Toward a Companion Tool for Social-Emotional Learning in Autism", Cambridge, MA: Massachusetts Institute of Technology Media Laboratory, Master of Science.

THOMAS, S. S., C. E. BUCKON, D. HELPER, N. TURNER, M. MOOR and J. I. KRAJBICH (2000), "Comparison of the Seattle Lite Foot and Genesis II Prosthetic Foot during Walking and Running", *Journal of Prosthetics and Orthotics* 12, pp. 9-14.

TORBURN, L., J. PERRY, E. AYYAPPA, and S. L. SHANFIELD (1990), "Below-knee amputee gait with dynamic elastic response prosthetic feet: a pilot study", *Journal of Rehabilitation Research and Development* 27: pp. 369-384.

WALSH, C., K. ENDO and H. HERR (2007), "A Quasi-Passive Leg Exoskeleton for Load-Carrying Augmentation", *International Journal of Humanoid Robotics*.

WALSH, C., K. PASCH, H. HERR (2006), "An Autonomous, Underactuated Exoskeleton for Load-Carrying Augmentation", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October 9-16.

WATERS, R. L. and S. MULROY (1999), "The Energy Expenditure of Normal and Pathologic Gait", *Gait Posture* 9, pp, 207-231.

WHITTLE, M. W. (1991), *Gait Analysis: an Introduction*, 3rd ed., Oxford: Butterworth-Heinemann.

WINTER, D. A. (1983), "Biomechanical Motor Pattern in Normal Walking", *Journal of Motor Behavior* 15(4), pp. 302- 330.

WINTER, D. A., and S. E. SIENKO, (1988),"Biomechanics of Below-Knee Amputee Gait", *Journal of Biomechanics* 21(5), pp. 361-367.

ZIEGLER-GRAHAM, K., E. J. MACKENZIE, P. L. EPHRAIM, T. G. TRAVISON and R. BROOKMEYER (2008), "Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050", *Archives of Physical Medical Rehabilitation* 89(3), pp. 422-429.

BBVA

The City to Come

Carlo Ratti
and Nashid Nabian
MIT SENSEable City Lab

What will the cities of tomorrow be like? Back in the 90s, many scholars speculated about the ongoing digital revolution's impact on cities, and the possibility of replacing physical space with virtual space, or atoms with bits. They fantasized about the dark, sexy image of disappearing urban spaces inhabited by individuals who would lead a mostly virtual life in cyber space, engaging in digitally encoded interactions rather than face-to-face communication.[1] Enthusiasts of digital technology pushed the envelope to the extreme by announcing the official death of history, space, time, geography, and cities, among other things.[2] The mainstream view was that digital media and the Internet would kill cities in the same way that they had killed distance. Technology writer George Gilder proclaimed that "cities are leftover baggage from the industrial era," and concluded that "we are headed for the death of cities" due to the continued growth of personal computing, telecommunications and distributed production (Peters and Gilder, 1995). At the same time, MIT Media Lab's Nicholas Negroponte wrote in *Being Digital* that "the post-information age will remove the limitations of geography. Digital living will include less and less dependence upon being in a specific place at a specific time, and the transmission of place itself will start to become possible" (Negroponte, 1995).

Yet, it became apparent in the years following the first wave of enthusiasm about digitality, that this was not the destiny of either our digitally enhanced race, or the constructed spaces and landscapes that accommodate our activities. Cities and the constructed spaces that they contain have been multiplying at an unprecedented rate, and the spatial production and consumption of mankind still fall very much within the physical realm. In fact, cities have never prospered as much as they have in the past couple of decades. For example, China is currently building more urban fabric than humanity has ever constructed in one era. And a particularly noteworthy moment occurred two years ago: for the first time in history, more than half the world's population—3.3 billion people—lived in urban areas.[3]

Thus, despite a generalized obsession with the vision of an all-digital world, a new situation has emerged where the digital and the physical world are merging, and atoms are augmented by bits of information. The

[1] There are numerous accounts of possible conditions of this nature. Sherry Turkle describes digitally savvy individuals for whom the worlds they inhabit through their computer screens are as real as the real world (Turkle, 1997).

[2] Vincent Mosco argues that these readings are caused by man's historical fascination with new technology. After examining the enthusiastic claims about end of space, time, history, economics, cities and the like by cyberspace, and going back in history to look at the similar mythic pronouncements prompted by past technological advances—the telephone, the radio, and television, among others—Mosco explains how such myths are created, and why we feel compelled to believe in them (Mosco, 2004).

[3] According to a UN report, "In 2008, the world reaches an invisible but momentous milestone: For the first time in history, more than half its human population, 3.3 billion people, will be living in urban areas. By 2030, this is expected to swell to almost 5 billion."

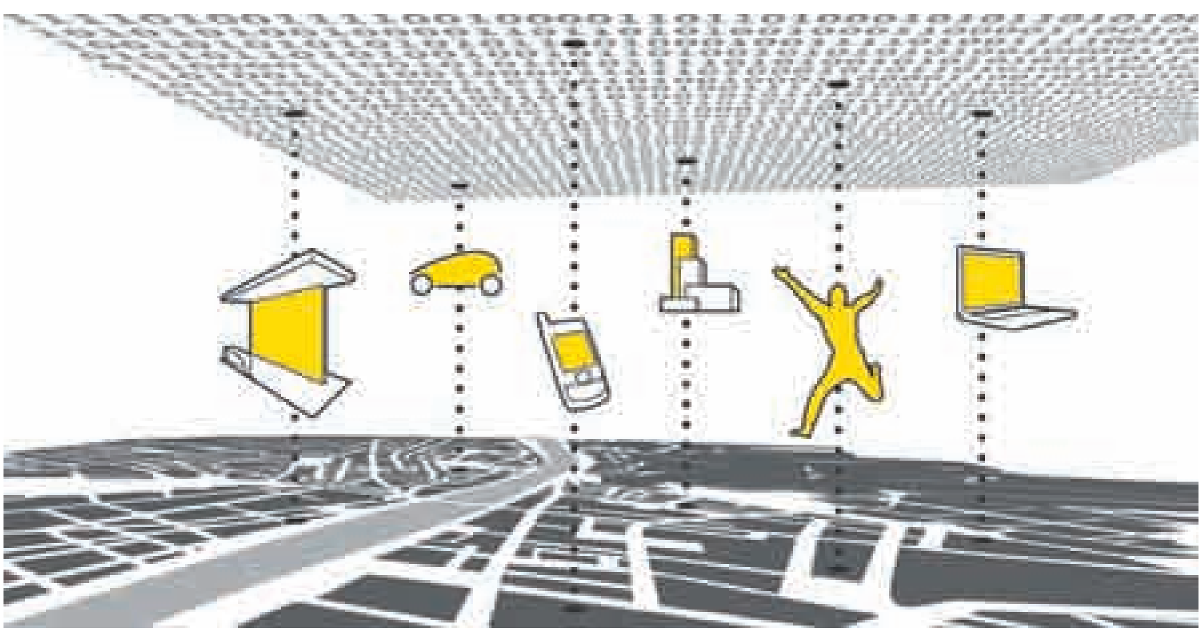
digital did not and will not kill the physical, as fantasized during the 90s. In fact, the digital and the physical are recombining, or, in the words of Hiroshi Ishii: "The bits and bricks are marrying."[4] A layer of networked digital elements blankets our built environments, blending the information sphere and the physical space inhabited by contemporary subjects in a seamless way.

What are the consequences of such transformations? This question can be addressed on several levels. In this article we will focus on one particular aspect that we have found most productive: the transformation of our cities into cybernetic, real-time control systems with a combined static and dynamic nature that consists of things that exist in the material sphere, and things that happen in the info-social sphere.

What follows is a speculation on the factors that will most significantly contribute to the birth of this new generation of urbanity. Our cities of the near future will operate as cybernetic systems that function via sentient control mechanisms. With a plethora of possibilities in telecommunication, people who live in digitally augmented cities will benefit from real-time access to vast repositories of information. And, with the aid of new sensing and actuating technologies, all constitutive elements of urban living will be transformed into context-aware, decision-making entities. In such intelligent environments, people will be able to be incorporated as entities with transient desires, needs and preferences: hyper-individualized "users" as opposed to generic "inhabitants."

We will conclude by speculating about the new generation of denizens that are to inhabit these cities: user-inhabitants who are digitally augmented and well informed about the dynamics of the cities that they inhabit. In other words, we will focus on how people who are environmentally digitally enhanced start acting like sensors in their

Figure 1. *How can a city perform as an open-source, real-time system?* MIT SENSEable City Lab Vision Poster



This illustrates my claim that cities and other human-constructed landscapes are and will be multiplying due to the demands of the increasing world population of urbanites (Ahmed Obaid, 2007).

[4] Hiroshi Ishii coined the term "tangible bits", which focuses on the idea of "graspable & manipulable" bits by "coupling the bits with everyday physical objects and architectural surfaces". Once bits of information become tangible, they can be considered building blocks of our inhabitable spaces, bridging the gap between cyberspace and physical space with digital technology. Thus, bits and bricks are married in a new paradigm shift in the realm of spatial practices (Ishii and Ullmer, 1997: 234-241).

[5] Mark Weiser is the father of *Ubiquitous Computing* or *Ubicom*. In his 1991 paper, "The Computer for the Twenty-First Century", Weiser discusses the idea of integrating computers seamlessly into the world: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable

> “With a plethora of possibilities in telecommunication, people who live in digitally augmented cities will benefit from real-time access to vast repositories of information”

own right, actuated in a real-time feedback mechanism between the city and themselves, and mediated with the aid of new digital technology and telecommunication networks, while the city itself is the interface for such mediation.

THE CITY AS A CYBERNETIC, REAL-TIME CONTROL MECHANISM

In his 1969 article “The Architectural Relevance of Cybernetics”, Gordon Pask proposed that architectural spaces should be designed as systems capable of responding to emerging conditions, and adapting to the needs of their inhabitants (Pask, 1969: 494-496). To this effect, he compared such spaces to cybernetic systems. Following the same line of thought, we may conceive of the digitally enhanced, postmodern city as a cybernetic mechanism that accommodates interaction in its capacity as a spatial system capable of extracting contextual information, acknowledging the inhabitants' desires and needs, and adopting behavior patterns based on what it learns.

Such a cybernetic urban system achieves its monitoring with sensing technology. It is conditioned through computational processes that are based on detected spatio-temporal changes. It is actuated through embedded virtual or physical agents that provoke changes detectable by the inhabitant, or that enhance the spatial experience of the occupant in an explicit or implicit way. It is also enhanced with memory of the past and anticipation of the future, and is endowed with some level of data connectivity, especially if the monitoring and actuating agents are physically separate and the distance needs to be technologically bridged. These steps imbue the city with a limited awareness of contextual change over time, and the capacity for responding to it accordingly.

In terms of sensing, cameras and microcontrollers are used ever more extensively to manage city infrastructure, optimize transportation, monitor the environment, and run security applications. Advances in microelectronics now make it possible to implement “smart dust” networks of tiny, wireless, micro electro–mechanical system (MEMS) sensors, robots or devices. Besides, we are witnessing an explosion in mobile-phone use around the globe. According to *ITU World Telecommunication Indicators Database,* more than four billion mobile phones were in use worldwide by early 2009. Across socioeconomic classes and five continents, mobile phones are ubiquitous: they allow us not only to communicate with each other in unprecedented ways, but to create a pervasive sensing network that covers the whole globe.

In terms of regulation and actuation, the city already contains actuators such as traffic lights, remotely updated street signage, etc. More profound actuation is relatively problematic: for instance, we cannot double the size of a street in real time if we detect traffic congestion. However, unlike other

from it.” He proposes the

real-time control systems, cities have a special feature: citizens. By receiving real-time information, appropriately visualized and disseminated, citizens themselves can become distributed, intelligent actuators who pursue their individual interests in co-operation and competition with others. Processing urban information captured in real-time and making it publicly accessible can enable people to make better decisions about the use of urban resources, mobility, and social interaction.

This feedback loop of digital sensing and processing could begin to influence various complex and dynamic aspects of the city, improving the economic, social, and environmental sustainability of the places we inhabit. Feedback loops could grow inside one another: buildings and other spatial devices throughout the city could become probes and ambient displays, but also evolve into real-time, responsive devices in their own right.

A cybernetic city operates on the logic of a hybrid computing paradigm that examines the ramifications of installing sensors that detect changes in the physical properties of the context; examines how an embedded microprocessor or computer processes the resulting digital signal; and finally examines how the system activates a series of actuators installed, embedded, or situated in the space. At times, the microprocessors act in isolation. Yet, once these micro-systems are networked, communication, sensing, and information-processing will disappear into the environment to create an *Internet of Things* with world-wide coverage, as in the idea of *Ubicom* proposed by Mark Weiser.[5] One consequence of this dual process of sensing and actuation within the contemporary city is particularly important: cities can start to work as real-time control systems, regulated by a number of feedback loops. In the text that follows, we would like to focus on the possibilities of sensing and actuation in the cities of the near future.

"Processing urban information captured in real-time and making it publicly accessible can enable people to make better decisions about the use of urban resources, mobility, and social interaction"

THE CYBERNETIC CITY AND ITS DIFFERENT MECHANISMS OF SENSING

In his *Traité des Sensations*, Étienne Bonnot de Condillac offers an interesting reflection on sensibility as the source of subjectivity.[6] He imagines a living statue, devoid of any sensations but the sense of smell. He walks the reader through a sequence where sensation leads to comparison, which leads to judgment, which leads to reflection and reasoning, which leads to abstraction: the sum of all of the above results in what can be characterized as understanding and [human] agency. This sensationalist approach can be applied to cities as cybernetic mechanisms. Sensors are devices capable of registering one or more quantifiable aspects in the contexts where they are used. Once populated with large number of sensors and endowed with the capacity to register changes in its context,

term "embodied virtuality" to refer to the process of drawing computers out of their discreet places, and seamlessly integrating computing into the environments of our day-to-day life experiences (Mark Weiser, 1991: 94-100).

[6] In *History of Philosophy,* Alfred Weber offers a very complete account and interpretation of the Condillac's sentient statue argument. I have borrowed from his interpretation in making my point (Weber, 1912: 399-403).

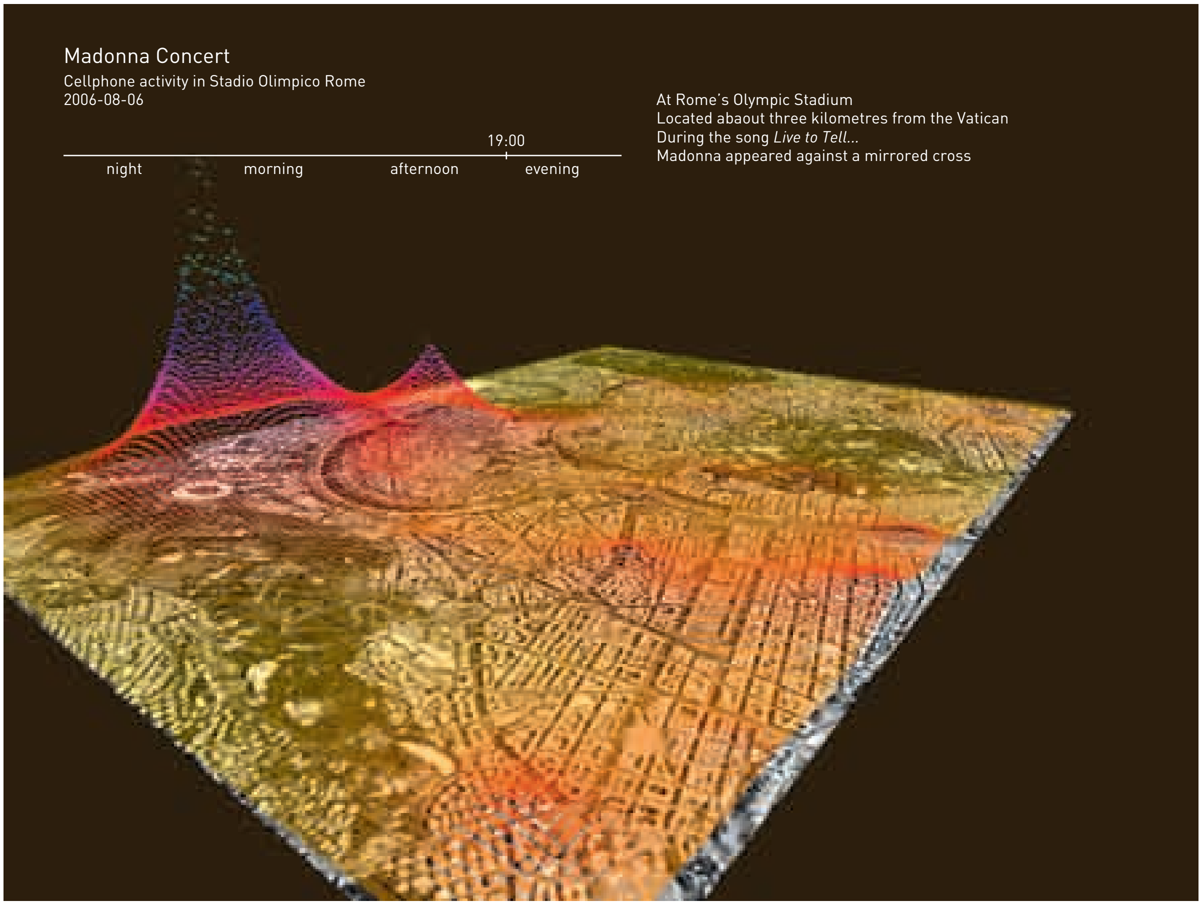
a city can acquire a limited level of agency through comparison, judgment, reflection, reasoning and abstraction.

To reach this goal, we should ask ourselves, how can we sense a city and its dynamics? One approach is to leverage systems already in place that have been developed for other reasons, but can function as a source of information on how our cities operate. A great example is the cellphone network. We define this as *viral sensing*, since the computational algorithms of such sensing practices install themselves on the digital networks that already augment cities, much like a virus settles in an already-operational environment within another organism in order to support its internal bio-processes. The premise of such sensing practices is that the contemporary subject voluntarily and involuntarily leaves digital traces on various networks that are juxtaposed over urban areas. Every time a credit card is used, a text message or an email is sent, a *Google* query is submitted, a phone call is made, a Facebook profile is updated, a photo is uploaded or tagged on Flickr, or a purchase is made on a major, on-line store like Amazon.com, an entry with the time and location of this action is added to a dataset on a central server, administered and maintained by the organizational entity providing the platform for these, and hundreds of other day-to-day operations. Once the datasets are spatially and temporally attached to entities and phenomena in the physical terrain, the urban landscapes that accommodate these traces are transformed to info-scapes. An info-scape, in this sense, is a digital terrain both temporally and spatially associated with the physical terrain. Info-scapes can be delivered on publicly used, architecturally installed digital screens, and personal, handheld computing units.

With advancements in the field of data connectivity and telecommunication technologies, connection to the aforementioned datasets of distant servers is improving, allowing data management engines to get real-time updates on the state of the monitored entities. The urban spaces of digitally annotated terrain augmented with mined datasets create mediated landscapes that allow for new forms of expression, such as public or museum exhibitions, or urban demos. For example, MIT SENSEable City Lab has conducted numerous experiments that led to such exhibitions and urban demos over recent years, some examples of which are *Wikicity Rome* and *NYTE*. In both cases, the real-time visualization of data mined from communication networks is cross-referenced with the geographical terrain, to allow for revealing urban dynamics in real-time to observers. Such technologically enhanced spectacles—real-time info-scapes projected onto architectural surfaces, or accessed via worn and handheld devices—provoke a temporary displacement of the observer from the physical terrain he/she inhabits to a distant location, providing him/her with a overview of the dynamics contained within the urban landscape.

The *Wikicity Rome* project tapped into aggregated data from cellphone usage. The resulting visualizations depicted the pulse points of the city, providing an overview of how the urban landscape is occupied, and where and in which temporal patterns the cellphone-using crowd is dispersed. Crowd sensing based on cellphone usage allows for spotting the hot locations and congested spots of the city in real-time. This can help administrative agencies to regulate traffic and the flow of resources within the city, based on real-time dynamics.

Figure 2. *Wikicity Rome*, 2007. MIT SENSEable City Lab: Assaf Biderman, Francesco Calabrese, Kirstian Kloeckl, Carlo Ratti, Bernd Resch y Andrea Vaccari



When the system was exhibited at the 10th International Architecture Exhibition of the Venice Biennale, researchers at MIT SENSEable City Lab also supplemented the resulting, cellphone-based evaluation of urban dynamics with data based on the instantaneous positioning of buses and taxis. This provided information about mobility, ranging from traffic conditions to the movements of pedestrians throughout the city, in real time. The visualizations provided a qualitative understanding of how the aggregated data of network cellphone usage and public transit locationing information can be used to provide valuable services to citizens and authorities. Researchers at MIT SENSEable City Lab believe that such information "can give city dwellers a deeper knowledge of urban dynamics and more control over their environment by allowing them to make decisions that are more informed about their surroundings, reducing the inefficiencies of present day urban systems" (Calabrese et al, 2010).

Aside from tapping into existing networks, customized sensor networks can also be implemented to decode various flows within the cities. The cybernetic city can receive its input from different networks of sensing mechanisms. The first is a network of centrally managed sensing agents, embedded within the fabric of the city. To this effect, smart dust saturates the space of urbanity, extracting large amounts of information about the processes contained within the built environment and constructed spaces, and channeling it to a central control and command mechanism. From there, this data is aggregated, managed, and used as the basis for how the monitored space of the city should be regulated and actuated.

The technology to geo-localize the whole surface of the world is not that far from becoming a reality. At this point, *Google Maps API* offers a two-dimensional, virtual model of the world. With advancements in the *Google Earth API* and platform, this virtual model is moving towards a fully realized, three-dimensional, virtual copy of the physical world. Imagine a day when such a model is augmented with geo-localized layers of information about each and every object it contains. This scenario would have two different but closely-related impacts on our consumption and production of space. First, the data extracted from all things would allow spatial practitioners to make sense of the dynamics of our constructed landscapes by analyzing these large, real-time data sets. This real-time knowledge of spatial dynamics could be fed back into the process of spatial design and the management of spatial resources. On the other hand, if access to such layers of information were democratized, then the inhabitants of our constructed landscapes would also benefit.

If a collectivity of sensors capable of communicating with a centrally managed server is embedded and distributed within a spatial context, the prospect of distributed sensing is shown as one aspect of a completely networked world, or an *Internet of Things*. If the sensors are well situated within the physical terrain, meaning that the digitally encoded transactions they send the server are annotated with their exact locations, this network allows a cybernetic system to create a virtual model of the physical terrain that is perpetually updated with real-time information about the context. Depending on the nature of the sensed data, the virtual model can represent different aspects of the physical terrain.

Sensor networks can consist of fixed sensor agents, or they can incorporate agents capable of navigating or probing the monitored terrain. If the sensors are embedded, they create a fixed backdrop for a virtual model of the monitored space's real-time dynamics that corresponds to the permanent geographical location of the sensors. If the nodes are implemented as dynamic probes, this backdrop layer is also animated. In some cases, once the geographic locations of the sensors are coded, the only transmission required to update the virtual model is the sensor-detected change, and the unique identification code of the sensor transmitting the data. In the second category of mobile sensors, each transmission also needs to be tagged or annotated by the current location of the sensor.

Furthermore, a sensor network may be implemented so that the agents automatically report the current state of the monitored phenomena at pre-defined intervals, or so that agents report their current state in response to an update request from the central server aggregating the sensed information. This gives us the dichotomy of self-reporting sensor networks versus request-based sensor networks.

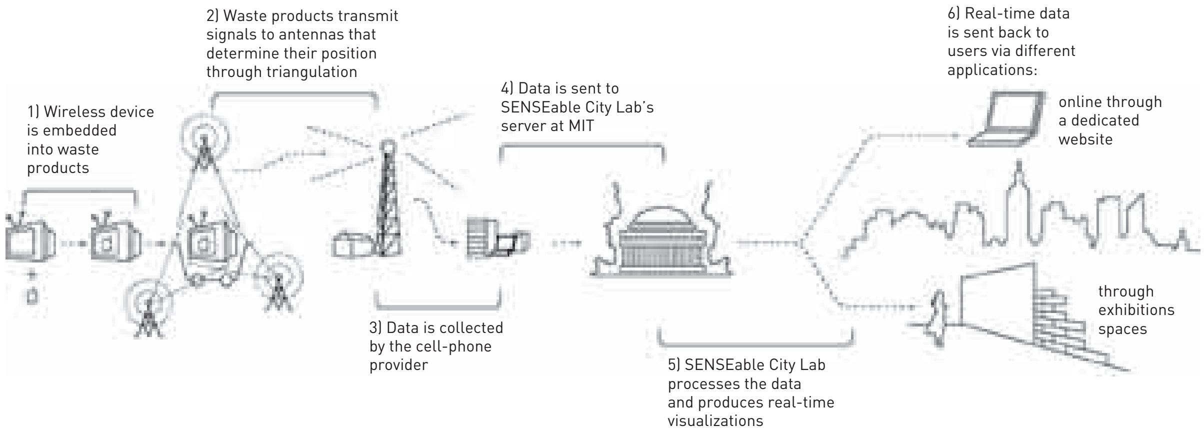
In urban sensing mechanisms that operate based on networks with central control-and-command structure, with algorithms that allow for the context-sensitive interpretation of the information transmitted to the database, we arrive at spatial scenarios where whole geographic regions are digitally augmented by the ubiquity of these sensor/transmitter devices. In the extreme version of this scenario, one can envision a world where any object is capable of sensing its context and reporting it back to a central data-management facility where its identity, its real-time location, and its contextual state are cross-referenced, stored, and managed computationally. A great example of such a scenario is the 2009 *Trash Track* project by MIT SENSEable City Lab.

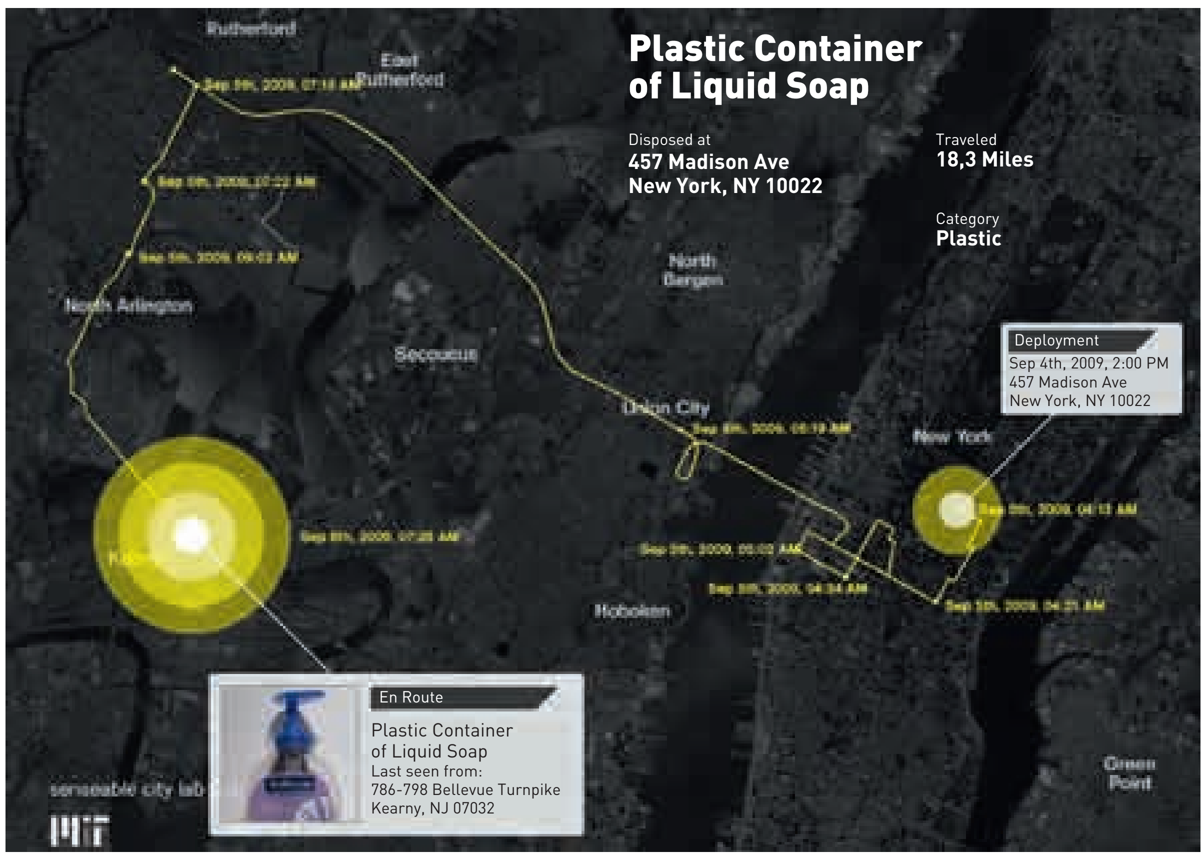
The project consisted of digitally enhanced tags that could be attached to objects and could report their location to an Internet backbone infrastructure via cellular network. *Trash Track* made use of these location-reporting tags to track urban disposal and study the efficiency of the urban waste-removal chain. The platform allowed designers and planners to analyze the acquired data, and make well-informed, high-level decisions about how the given constructed landscape is managed. Therefore, a multiplicity of questions about the dynamics of the urban removal chain could be addressed empirically: Is our removal chain efficient? Is hazardous waste managed properly, or are there loopholes in our system that need to be taken care of? Is the recycled waste really recycled, or does it end up in dumps? The *Trash Track* system can have a great impact on the nature of the perceptual relationship that a city or region develops with its waste disposal habits.

Generally, people assume that once they dispose of waste, it is no longer their responsibility. Offering a real-time view of how the disposed items travel through the landscape of their daily lives will expand each citizen's perceived sphere of responsibility from the domestic space, to the space of the city. Perhaps such real-time urbanity can result in a more responsible urbanity. Yet, *Smart Trash* is but one possible scenario in a more comprehensive conception of a world populated with sensors.

As previously discussed, in viral sensing, the involuntary digital footprints of contemporary urbanites stored in central databases of service providers is the basis for making sense of city dynamics. Again, sensor networks employ a top-down architecture where all the sensors report information from the environment to a central database, from where this information is aggregated, managed, and stored. Instead of such top-down approaches, we should also consider more grassroots, bottom-up systems for sensing the dynamics of cities. One possibility is thinking of each urbanite as a *human sensor*, an agent for sensing and reporting on his or her individual experience through tapping into data generated by user-contributed content on content-sharing platforms. Hence, we arrive at the third possibility of urban sensing: *crowd-sourcing*. User-generated content-sharing platforms allow everybody to report his or her experience to others in real-time, and in a multi-modal, high-resolution format. On these platforms, the users constantly project the physical world onto the digital world. Websites such as Flickr, Twitter, Facebook, and Wikipedia are repositories of what people "sense" in the city. This dynamic is gradually creating a digital world that mirrors the physical world. For almost every city in the world, a parallel digital version, as rich in diversity and content as its physical counterpart, is spread across different platforms and systems. This is because the digitally- augmented, contemporary urbanite uploads pictures of popular events, sends tweets about new happenings in real time, and creates and updates pages on Wikipedia about the city. These acts of communication generate different kinds of data that provide

Figure 3. *Trash Track*, 2009. MIT SENSEable City Lab: Carlo Ratti, Director; Assaf Biderman, Assoc. Director; Dietmar Offenhuber, Team Leader; Eugenio Morello, Team Leader; Musstanser Tinauli, Team Leader; Kristian Kloeckl, Team Leader; Lewis Girod, Engineering; Jennifer Dunnam, E Roon Kang, Kevin Nattinger, Avid Boustani, David Lee, Programming; Alan Anderson, Clio Andris, Carnaven Chiu, Chris Chung, Lorenzo Davolli, Kathryn Dineen, Natalia Duque Ciceri, Samantha Earl, Sarabjit Kaur, Sarah Neilson, Giovanni de Niederhausern, Jill Passano, Elizabeth Ramaccia, Renato Rinaldi, Francisca Rojas, Luis Sirota, Malima Wolf, Armin Linke, Video





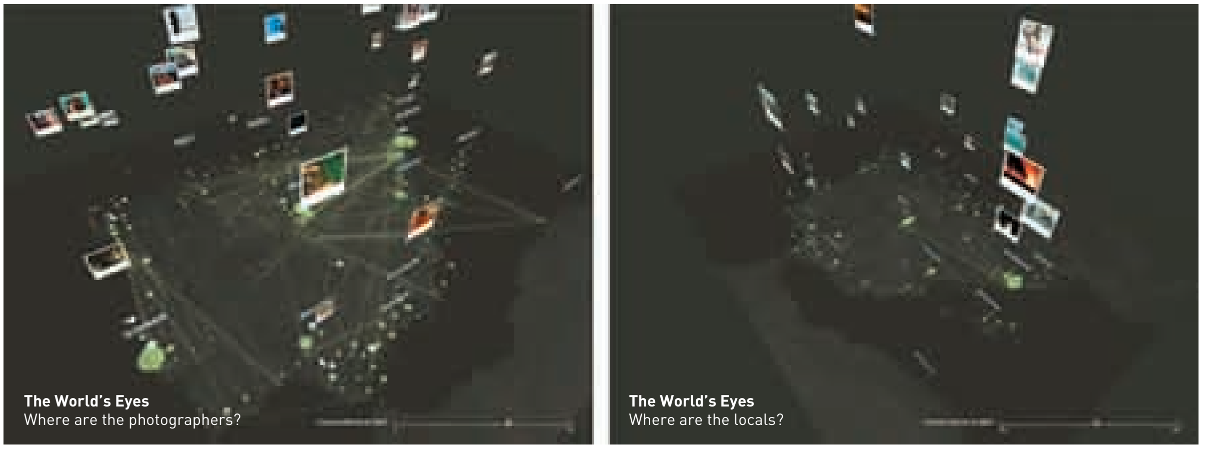
unique views on how people experience, navigate, and view the city. The crowd therefore becomes a distributed network of sensors that allows us to understand the dynamic patterns of the city and the experiences of its citizens at a quasi real-time rate. Hence, we call this phenomenon *crowd-sensing* (Pereira *et al.*).

Accessing the possibilities offered by user-generated content-sharing platforms, researchers at SENSeable City Lab have conducted various projects that focus on revealing the dynamic of civic landscapes, as viewed and collaboratively reported by their inhabitants. For example, in the *Los Ojos del Mondo/The World's Eyes* project, the attractiveness and popularity of places and events are revealed by visualizing the density of user-generated data, in particular, the photographs tagged with information about their location and time uploaded by Flickr users. Then, user-generated electronic trails based on the sequences of photographs are used to reveal the presence and movement of visitors in a city. Such data visualizations that geo-localize the content generated by the user's experience of a given urbanity reveal how cities are interpreted by their occupants—e.g., which locations are considered more or less important, and what is captured by the eyes of the people who are there. The "virtual city" created through the geo-localization of user-generated content reflects reality of the city dynamics, and can become a powerful resource for understanding these (Pereira *et al.*).

For example, an animation of the photos geo-tagged to different neighborhoods of Barcelona with descriptive tags that relate to "partying" in the summer of 2007, shows that Barcelona's old town (Ciutat Vella) is where one goes to have fun. This observation is validated by the fact that the area contains both a high density of tourists, the Bohemian district of Gracia, and the Forum area (where various music festivals are held).

Another visualization in the same set looks at how Spain is photographed by tourists over the course of one year. While the photos overlap in certain locations and expose places that attract the photographer's gaze, in other locations, the absence of images is eye-catching, revealing the more introverted parts of Spain.

Figure 4. *Los Ojos del Mondo/The World's Eyes*, 2009. MIT SENSEable City Lab: Carlo Ratti, lab director; Assaf Bidermann, associate director; Fabien Girardin, project leader; David Lu, visual designer; Andrea Vaccari, data mining. Universitat Pompeu Fabra: Ernesto Arroyo, interaction designer.



> "Actuators and information delivery interfaces are the components of the space controlled by the output of the operating system, based on changes registered by the sensors and reported to the operating system as an input"

Now that we have covered three different ways of providing input for the real-time control mechanism of a cybernetic city—viral sensing, implemented sensor networks, and crowd-sensing—we will move on to speculating about the output of such urban systems, or in other words, the actuation mechanisms.

THE CYBERNETIC CITY'S DIFFERENT MECHANISMS OF ACTUATION AND REGULATION

Parallel to mechanisms of sensing, mechanisms of actuating are integral to cities that are cybernetic systems. In terms of spatial actuation and regulation, we can speculate on two sets of possibilities. The first is regulating the landscape through actuator agents embedded within the space and controlled via algorithms that are conditioned by the information received from various sensing mechanisms. This vision opens up a multiplicity of possibilities for the design and implementation of responsive environments and interactive spaces by integrating digital technologies into the design of buildings and artifacts.

Actuators and information delivery interfaces are the components of the space controlled by the output of the operating system, based on changes registered by the sensors and reported to the operating system as an input. The actuation of a digitally augmented space can be considered in terms of provoking the type of physical motion we see in kinetic architectures. Motion-initiating agents can provoke rotation, vertical and horizontal disposition, or vibration in the elements of spatial settings that are of a substantive nature. Another possibility is to use materials that change shape when exposed to an electrical current. For example, depending on the pattern in which muscle wires are woven in to the fabric of an architectural surface, when an electrical current is applied, the surface changes form to accommodate the change in the length of the wires.

Once the inhabitable spaces of cities are transformed into context-aware, decision-making entities with the aid of sensing mechanisms that are also capable of analyzing the sensed data, the human subject inhabiting the space can be incorporated as an entity with transient desires, needs and preferences. This allows the environment to acknowledge its inhabitants' input, or at least the specificities of their behavior. Inhabitants are then identifiable, each deserving specific treatment from the space he/she inhabits. A user-subject is a hyper-individualized inhabitant, and an interactive space respects the specificities of, and offers a customized experience for each one.

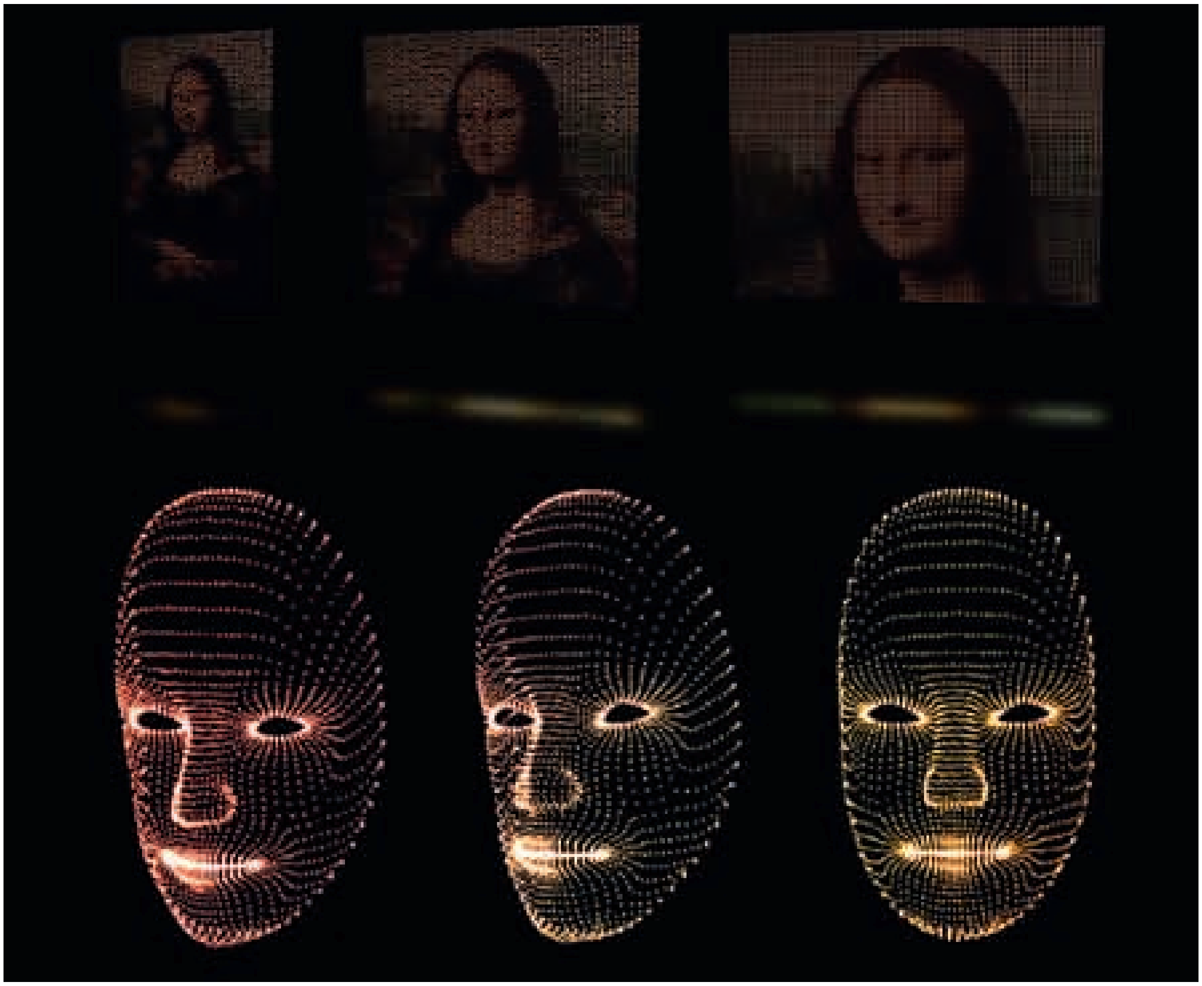
Physical locomotion is not the only way of initiating change within spatial settings. Change may also be invoked by manipulating the soundscape to inject the space with

Figure 5. the Cloud, 2009. Architecture by: Carlo Ratti, Walter Nicolino, Alex Haw. Team from carlorattiassociati and MIT SENSEable City Lab: Giovanni De Niederhausern, Alberto Bottero, Pietro Leoni, Coen Smets, Assaf Biderman, Mauro Martino, E Roon Kang.

auditable [white] noise, or melodic and musical pieces that vary over time in terms of content, volume, and intensity. Perceivable change can be suggested visually, by light-emitting agents that vary in terms of intensity and color. Dispensing fog in different densities can also manipulate the visual aspects of the space by changing the depth of the perspectival field. Architectural surfaces can be endowed with variable visual characteristics, such as transparency or color, based on the amount of electrical current to which they are exposed. Furthermore, change may be evoked through digital screens embedded in architectural surfaces that deliver animated imagery to the subjects inhabiting the space. This is the case for the *CLOUD* display system. The patterns of its animated, spherical skins offer a civic-scale interface for delivery of real-time information to the inhabitants and visitors of the city.

Another example of display technologies that can actuate the space of the city is explored in the Flyfire project. Flyfire, a project by the SENSEable City Laboratory in collaboration with ARES Lab (Aerospace Robotics and Embedded Systems Laboratory), uses a large number of self-organizing micro-helicopters that contain small LEDs and act as smart pixels. The helicopters are controlled to create synchronized motions, and form elastic display surfaces. This allows for the transformation of any ordinary space into a highly immersive and interactive display environment. The proposed mechanism explores the possibility of a free-form spatial

Figure 6. Flyfire, 2010 (Top: Raster Image Display, Bottom: Vector Image Display). SENSEable City Lab: Carlo Ratti, Director; Assaf Biderman, Assoc. Director; Carnaven Chiu, Team Lead & Visuals; E. Roon Kang, Team Lead (2nd Phase); Caitlin Zacharias, Shaocong Zhou. ARES Lab: Emilio Frazzoli, Director; Erich Mueller, Engineering

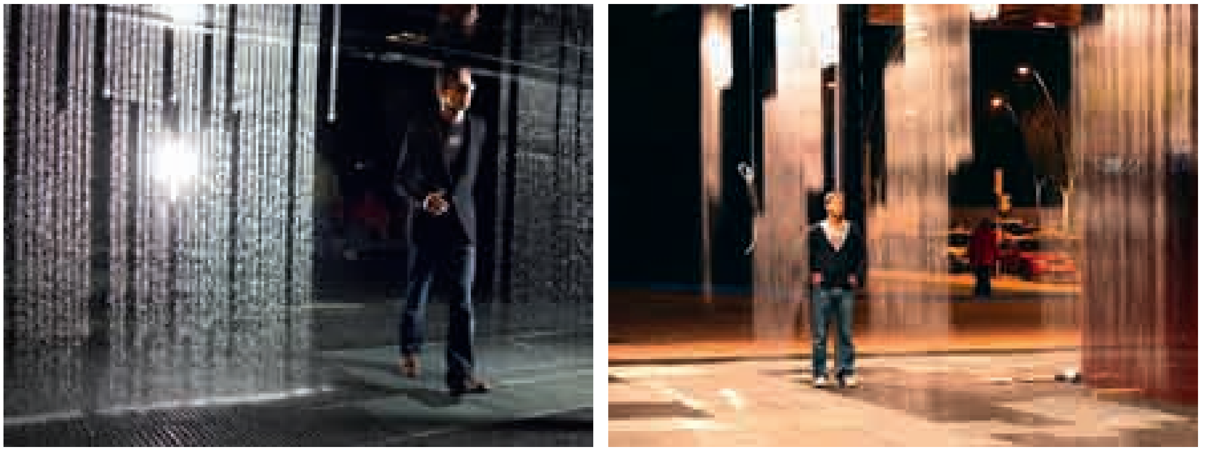


display that consists of a swarm of pixels that self-organize in real-time to adapt to the display requirements of any given scenario.

Space can also be suggestively manipulated in terms of thermo-ception. Actuators can provoke change in the environment through heating and cooling, or humidifying and dehumidifying mechanisms. Thermo-ceptory aspects of space can be conditioned using actuators such as fans that manipulate the pattern and intensity of air-flow through the space. Alternately, change can be expressed in olfactory terms by embedded, scent-emitting actuators that disperse patterns of odorous gases and liquids. In the extreme case, embedded liquid or gas dispensers will allow the designer to envision architectural settings that are of a less substantive or material nature, which is the case for the *Digital Water Pavilion* project.

Yet, manipulating space through embedded actuators is not the only possible means of spatially regulating cybernetic urban systems. The inhabitants of the cities themselves can be considered possible agents of regulation and actuation. From

Figure 7. *Digital Water Pavilion*, 2008. Architectural design: Carlorattiassociati. Landscape architecture: Agence Ter. Graphic design: Studio FM Milano. Engineering: Arup. Interactive water wall engineering: Lumiartecnia International. Lead contractor: Siemens. Site supervision: Typsa. Expo gateway preliminary design: MIT Senseable City Lab. Digital Mile design: MIT Department of Urban Studies and Planning / City Design and Development Group. Interactive water wall concept: MIT Media Laboratory / Smart Cities Group



this point of view, the space of the city is envisioned as the provider of real-time access to information for a body that corporeally inhabits it. Hence, spatial design does not limit itself to the allocation of material resources, but takes into account the temporal allocation of information relevant to the specific location or context of those occupying it. The new analytical understanding of the spatial dynamics offered by the delivery of real-time and geographically situated information can be fed back to the individuals inhabiting these landscapes, to help them make well-informed decisions. An example of this approach is the real-time, context-sensitive service offered by cellular networks that assess crowd density based on cellphone usage in an area, and deliver this information to city residents who wish to identify popular "hot spots." In such scenarios, not the space but the inhabitants of the space are actuated, and efficient regulation of spatial dynamics is based on their decisions.

This is what we see as the most promising characteristic of the city of the future: the fact that it is made "smart" by the collaborative activity of its citizens. The citizens have the potential to function as sentient, self-reporting agents, contributing to monitoring the city as a cybernetic organism. On the other hand, they can be actuated and their actions can be self-regulated based on real-time knowledge of the city dynamics, conveyed on information-delivery platforms.

Such a city will be a desirable place in which to live and work, in that it offers a platform for reinforcing identity and culture through collaboration. Collaboration in defining and re-defining the ephemeral dimension of the city can be viewed as one way of making it more engaging, and making its citizens more invested in it. A city that is open to individual modifications enables people to imprint intentional traces of themselves upon it. In this way, the city itself becomes a limitless canvas of collaboration, resulting in a constant feeding of human input.

A city whose inhabitants become sensors, and which is actuated by the results of the real-time information provided to them about its internal dynamics, will be more responsive to concerns about adaptability, efficiency, and optimal operation. Therefore, although augmented cities respond to concerns about function, structural durability, and aesthetic desirability, the focus of designing such spaces will inevitably shift to the issue of performance. After all, any space capable of self-adapting to new conditions is not there to merely endure, but to "perform" with efficiency. In the end, digitally augmented cities are performing cities, and given the right technology, the limit to how well they can perform is the limit of the imagination and the desire of those who design and inhabit them. To this effect, the cybernetic city will function as a medium through which subjects within the space communicate amongst themselves, transforming them from passive inhabitants to active participants in, or actuators of spatial scenarios. Hence, we end our consideration of the city of the future with the terms "user-participant-inhabitants" or "sensor-actuator citizens" to refer to the people who are its citizens, and who, we imagine, will be ultimately responsible for the cybernetic organism they inhabit.

REFERENCES

Ahmed Obaid, T. (2010), "Peering into the Dawn of an Urban Millennium", *State of World Population 2007* (New York, NY: United Nations Population Fund, 2007), http://www.unfpa.org/swp/2007/presskit/pdf/sowp2007_eng.pdf (last accessed April 12, 2010).

Calabrese, F., M. Colonna, P. Lovisolo, P. D. Parata and C. Ratti (2010), "Real-Time Urban Monitoring Using Cellular Phones: a Case-Study in Rome", *IEEE Transactions on Intelligent Transportation Systems* (2010).

Ishii, H. and B. Ullmer (1997), "Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms", in S. Pemberton (ed.) *Proceedings of the ACM CHI 97 Human Factors in Computing Systems Conference*, New York, NY: ACM, 1997, pp. 234-241.

Mosco, V. (2004), *The Digital Sublime: Myth, Power, and Cyberspace,* Cambridge, MA: The MIT Press.

Negroponte, N. (1995), *Being Digital*, New York: Knopf.

Pask, Gordon (1969), "The Architectural Relevance of Cybernetics", *Architectural Design 39*, pp. 494-496.

Pereira, F. C., A. Vaccari, F. Giardin, C. Chiu and C. Ratti, "Crowdsensing in (a?) Web: Analyzing the Citizens' Experience in the Urban Space," SENSEable City Lab, MIT, USA.

Peters, T., and G. Gilder (1995), "City vs. Country: Tom Peters & George Gilder Debate the Impact of Technology on Location", *Forbes ASAP Technology Issue*, February 27, 1995, p. 56.

Turkle, S. (1997), *Life on the Screen: Identity in the Age of the Internet,* New York: Simon & Schuster.

Weber, A. (1912), *History of Philosophy*, trans. Frank Thilly, New York: Charles Scribners Sons, pp. 399-403.

Weiser, M. "The Computer for the Twenty-First Century," *Scientific American,* September 1991, pp. 94-100.

Future Opera for Robots and People Too

Tod Machover
MIT Media Lab

We all love music deeply; opera adds every other conceivable form to augment and unify the senses *around* music. Music and opera are capable of entertaining, stimulating, moving and transforming us as few other activities are. In fact, there is increasing research on the how and why of music's power, some of it—like the much-hyped 'Mozart Effect'—suggesting that merely listening to music on your iPod while driving, reading, sleeping or perhaps even playing music to your baby in the womb is enough to let music work its full magic.

Unfortunately, that isn't quite true. Music exerts its power when we are actively engaged, not when we listen subliminally. For this reason, I have been working with my group at the MIT Media Lab to create musical tools—often with specially designed technologies—that enable everyone to participate directly in music-making regardless of background.

This field has undergone a revolution in the past several years through the huge public success of the *Guitar Hero* and *Rock Band* videogames. Alex Rigopulos and Eran Egozy, students of mine from the MIT Media Lab, developed them based on ideas that we were working on in the early 1990s. The good news about *Guitar Hero* and *Rock Band* is that they clearly demonstrate the public's willingness to dive in and immerse themselves in music-making, given the right environment. The bad news is that neither platform is truly musical, nor do they encourage learning, expression or creativity.

What would happen if we could combine the excitement and 'stickiness' (bordering on addiction) of *Guitar Hero* with a more sustaining, personal and open-ended musical experience? How could we embed such a new type of activity in a more integrated musical ecology, where the current exaggerated distinctions between celebrities and amateurs would be diminished and the level of musical sophistication, excellence and, hence, enjoyment would be raised for all?

The research projects that I direct at the MIT Media Lab and many of the musical projects I undertake are attempting to nudge the field in these directions. We started by developing Hyperinstruments for some of the world's greatest performers, including Yo-Yo Ma and Prince, as well as orchestras, chamber music ensembles and rock bands.

All kinds of sensors are built into the Hyperinstrument so it knows how it is being played. By changing the interpretation and feeling during performance, a cello—for instance—can be morphed into a voice or a whole orchestra or something that nobody has heard before.

As these virtuosic Hyperinstruments developed, we started imagining that we could use similar technologies and interpretation strategies to produce instruments and interfaces for music-lovers who weren't highly trained virtuosi. We have designed a series of such instruments and interfaces. One of the largest collections was the Brain Opera that was launched at the first Lincoln Center Festival in New York in 1996, toured the world and is now permanently installed at the Haus der Musik in Vienna.

Tod Machover and Yo-Yo Ma playing the Hypercello

We created a rather large orchestra of especially designed fantasy instruments (including Rhythm Trees, Harmonic Driving, Gesture Walls and Melody Easels) so that anybody could play them using natural skill. You can play a video game, drive through a piece of music, use gestures to control huge masses of sound, touch a special surface to make melodies and use your voice to make a whole aura. We designed the Brain Opera for adult concert-goers, but found that everywhere we went it was most easily understood and most creatively manipulated by the youngest (under 8) and oldest (over 70) visitors. This was perhaps due to lack of inhibition and desire for social play and creativity among those ages. We therefore decided to concentrate on activities that might engage those groups more directly.

This led to our Toy Symphony project (2002–2005), which attempts to reconsider how to introduce children to music in the most immersive, creative and enjoyable way possible. The goal is to have kids fall in love with making music first and then demand to learn more because of that love. We designed a set of new Music Toys, including the soft, squeezable Music Shapers that manipulate intensity and tone colour; Beatbugs, which capture rhythms that can be manipulated and shared with friends; and a software-composing environment called Hyperscore that lets anyone compose original music by shaping lines and colours. Another goal of Toy Symphony was to develop a project model—learning musical skills, creating new music and then rehearsing and performing a concert—that would bring children and orchestras together.

Our results with Toy Symphony were encouraging enough to make us decide to bring this model to other populations where there might be clear impediments to personal expression and creativity, where music, made accessible through new musical tools, might be an ideal medium. In 2004, we

Brain Opera at the Haus Der Musik in Viena

began concentrating on providing musical experiences and tools—based on ongoing research by colleagues at the MIT Media Lab and increasingly around the world—to help improve health, diagnose illness and provide a medium of expression and communication that would otherwise be lacking. This new area of Music, Mind and Health has led to research in using music for early detection of Alzheimer's disease, for social and emotional adaptation for autistics, for aiding physical and mental rehabilitation and for a growing number of other areas. With my student Adam Boulanger, I started this work at Tewksbury Hospital near Boston, where we were invited to work with a group of long-term residents with a wide range of severe physical and mental disabilities. We organized composing workshops with Hyperscore that resulted in a series of public concerts featuring music by patients. This process has become so successful that it has been replicated at many sites, resulting in marked and unexpected improvement in a wide range of conditions, and inspiring a number of patients to themselves mentor others in the uses of new tools and environments for creative musical expression.

A performance system designed for and with Dan Ellsey is an example of a new category of interfaces that we call Personal Instruments. Even an instrument as sophisticated as the Hypercello we designed for Yo-Yo Ma is a generalized instrument. In other words, anyone familiar with cello technique can play it, drawing more from it according to one's mastery and understanding. But Dan's instrument was designed for him and him only: it takes account of his particular style and detail of moving and the way that translates into musical expression, and compensates for his particular physical limitations. Future instruments and interfaces can and must be adaptable and tunable to each of our skills and limitations. For us, Dan's performance system represents the first step in this direction.

How did I come to undertake such unusual work? Through a desire to compose music, the activity that I love to do most. It is what best combines my various skills and interests—imagination, reflection, organization and the desire to communicate my thoughts and emotions to anyone who will listen. I also love solitude: I do my creative work in an 18th-century barn on our farm near Boston, where I can pursue my ideas without the need to explain or translate until all is ripe and ready. So it may seem like a paradox that another large chunk of my life is spent in one of the world's most futuristic, collaborative and intensive centers of technological invention—the MIT Media Lab. But the attractions and complexities of merging these worlds are central to how and why I work, and grow from seeds planted when I was very young. My mother is a Juilliard-trained pianist and a remarkable pedagogue and my father is one of the pioneers of computer graphics, but it actually took me a while to start combining these fields. I grew up as a cellist, first playing solo Bach, then chamber music (I never particularly enjoyed playing in orchestras), and then, by high school, original composed or improvised music using a wired and transformed rock cello that I created by placing large headphones around the cello for amplification, then sending the sound through tape recorder loops and analog transformation processes.

The appearance of the Beatles' *Sgt. Pepper's Lonely Hearts Club Band* had changed my life: it suggested a music that ideally balanced complexity and directness. There was a downside, though: as a product of the recording studio, most of the Beatles'

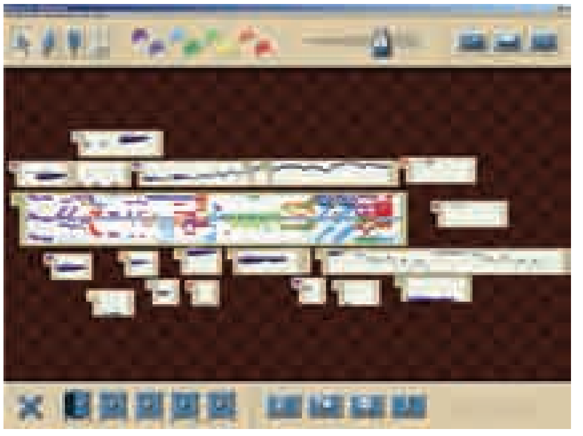
Music Shaper of Toy Symphony

music after 1967 couldn't actually be played live. That's when I started imagining a performance mode that would combine the physicality and intimacy of solo cello and the unhinged creativity of the recording studio. I was driven by the urge to bring this strange, enticing and intricate music filling my head out through my arms and fingers and into the world.

This desire compelled me not only to compose the music I was imagining, but also to invent new instruments and new modes of playing them, something that I never thought as a kid that I'd end up doing. So along with my colleagues and students at the MIT Media Lab I designed the projects outlined above.

Inventions like these have been part of a trend that has yielded amazing developments over the past 10 years. Technology has democratized music in ways that are surprising even to me, revolutionizing access to any music anytime with iPod and iTunes, opening interactive music-making to amateurs with *Guitar Hero* and *Rock Band*, providing digital production and recording facilities on any laptop that surpass what the Beatles used at Abbey Road, and redefining the performance ensemble with initiatives like the Stanford University iPhone Orchestra and YouTube Symphony.

Hyperscore



In fact, near the end of 2010 one wonders whether there is any more music technology to invent, or whether our musical imaginations and artistic cultures simply need to catch up. The answer is both, and then some.

For the first time in my career, I feel as if there are enough tools on my laptop, enough brilliant and inventive playing chops amongst the younger generation of performers, enough ooomph in the iPhone, and increasing openness and entrepreneurship in musical organizations both large and small to stimulate my imagination and allow for the production and dissemination of my somewhat unusual creations.

But even though these evolving music technologies are already very powerful and increasingly ubiquitous, we can also see their current limitations and potential risks. *Guitar Hero* is rhythmically exciting but not yet expressive or creative enough—a "sticky" but not "open-ended" experience that does *not* obviously lead to better musicality, listening or ensemble awareness. The iPhone is a remarkable little chameleon but lacks the touch and sensitivity of even the simplest traditional instrument, better for selecting and switching than for subtly shaping. Amplified sound is loudly present and "surrounds" us ever more, but still emphasizes the boom box aspect rather than the "still small voice." And there isn't yet a performance measurement system that could come close to interpreting the exuberance, range and immediacy of someone like conductor Gustavo Dudamel or truly enhancing the experience of an "unplugged" symphony orchestra.

As a composer, I find that each new piece I undertake suggests exciting but daunting technological challenges; my imagination just seems to be wired that way. My current project, the opera *Death and the Powers*, is one example.

I had been invited to imagine a new (and unusual) opera by the Opera of Monte Carlo, and two fundamental impressions came to mind early on. The first came from thoughts about mortality and how difficult it is to sum up one's life in a way that can be shared and transmitted to loved ones through generations, and how music has a particularly powerful capacity for collecting and concentrating multiple experiences, then burning them indelibly into our memories. And I started imagining that this web of musical memories—the embodiment of an entire life—needed to transcend traditional notes and instruments, jump off the stage and physically envelope the listener, both aurally and visually. This turned into a mental impression of floating, undulating, palpable 3-D sounds represented visually through slowly moving, morphing objects filling a stage—like *Fantasia* become physical (but with *my* music and *without* dancing elephants). I felt the need to go beyond the flatness and harshness of usual multimedia tools to create something that was at the same time transcendent and magical but also completely human and down-to-earth.

I then sought out collaborators—the poet Robert Pinsky and the playwright Randy Weiner—to turn these initial impressions into an opera, a form that has long attracted me for its use of word and image to ground music's abstract qualities in concrete human experience. Together we crafted a story about a man who longs to leave the world in order to pass to a higher level of existence, but wants everything about himself—his memories, his ability to influence others, his contact with those he loves, his legacy—to remain behind.

This story evolved into a full opera libretto in which the main character, named Simon Powers, switches on The System at the end of Scene 1: he becomes embodied more and more in his surroundings, forcing those left behind to decide how to communicate with him or it, whether to follow, and what part of his legacy to retain or reject. The stage itself becomes the main character in the opera, taking over from—and extending—the physical presence of the singer. Realizing this vision was a daunting challenge, but happily, with the collaboration of the director Diane Paulus, the designer Alex McDowell, the choreographer Karole Armitage and my group at the MIT Media Lab, we designed sighing walls, morphing furniture, gliding robots and even a resonating chandelier to create The System on stage—and to make it "sing."

In helping to tell this story and to sonify the score, all aspects of this physical set translate and amplify Simon Powers' human presence, using our new technique of Disembodied Performance, challenging the current limits of our ability to measure and interpret all the subtleties of a great performance. The techniques we developed yield surprising results, turning elegantly refined gestures, barely perceptible touch, and the gentlest breath into sounds, shapes and movements that convey personality and feeling without looking or sounding exactly like a human being, although we end up feeling extremely close to Simon. This whole infrastructure is a new kind of instrument, and we indeed learned how to "play" it in time for the world premiere performances of *Death and the Powers* in Monaco in September 2010. All of the unusual elements of this massive project—music, story, words, visuals, movement, robotics and more—came together and appeared to be more than the sum of their parts. An uneasy relationship was established—as desired—that invited audiences to question the boundary between humans and machines, and often established an emotional connection with the chorus of OperaBots.

These OperaBots frame the opera by accepting to perform this inherited story, left to them in a future time when there are no more humans on Earth. Once the opera starts, the OperaBots are almost always on stage, reacting to live performers, commenting on the action, being the playthings or "pets" of Nicholas, Simon Powers' assistant who built them and being sort of intermediaries between the humans and The System. They are not exactly individual characters, but they do have individualized choreographies and behaviors, gliding and twisting about, flashing and modulating light and—indeed—singing from time to time. The OperaBots have "character"—they are fun, interesting, engaged, energetic—but they do not understand the kinds of questions that give meaning and texture to human lives: relationships, time, touch, sacrifice. They care about the actions of the human characters, but they do not have the kind of motivations that underlie Simon Powers' final

The tenor Hal Cazalet in *Death and the Powers*; the Operabots in the background

confrontation with his daughter Miranda, where he pleads with her to enter The System with him, and she must decide what she would gain or lose by doing so.

Works of art do not have one single point or message. But two underlying inspirations behind this project were, first, how to allow technology to enhance human presence and communication on stage, as opposed to the huge distancing that happens more and more in mega-spectacle rock concerts where ugly, loud sound is pushed from the stage and performers look like ants against giant TV screens.

And I wanted to explore the possibility and poignancy of what is easy and what is hard to communicate between any two people—and especially across generations. I also wanted to create a journey where these questions and feelings would come alive through memorable melodies, unusual sonic textures, and pulsating rhythms…with the help of a few robots.

I believe that *Death and the Powers* is innovative in quite a few ways. An underlying goal of this opera has been to create a form of live performance which goes well beyond the typical multimedia practice of contemporary performance as seen constantly in—for example—arena rock shows. The norm these days is to create painfully overpowering sound to fill huge arenas, and gigantic video displays that dwarf human performers (even in U2's most recent tours, for example) rather than enhancing human presence. That is why we have instead created a stage filled with *animated physical objects*, from musical robots to animatronic walls to the Musical Chandelier. All translate sophisticated technology into physical form (including one of the most sophisticated sound systems ever used for live performance) with the goal of connecting audiences to the human beings on stage.

In addition, we have pushed the boundaries of robotic performance, not just with our autonomous OperaBots, moving walls of The System, and Chandelier, but with furniture robots that imbue inanimate objects with believable human characteristics. The opera might definitely induce people to think much more broadly about the potential of robots, and the productive relationship between robots and humans.

In order to control the unprecedented complexity of robots, visuals and sounds, we have developed special software to allow intermedia, integrated design of all aspects of the show, sophisticated enough for the most demanding AI programmer to fine-tune, but intuitive enough for a non-tech director or choreographer to feel comfortable with. This same software allows for real-time performance control of every aspect of the show, with perfect event synchronization and subtle inter-reaction between elements, so that everything on stage becomes part of a single, integrated "system."

And we have developed new sensing and interpretation technologies so that a combination of the singer's conscious (voice and hand gestures) and unconscious (breathing, heartbeat, muscle tension, etc.) behavior can be translated to control all of the interconnected stage elements, so that the stage and set itself truly feel as if they *are alive*, creating an uncannily believable representation of the human being who is no longer present. We believe that this technology could have significant impact on the future of telepresence and expressive communication and collaboration over distance.

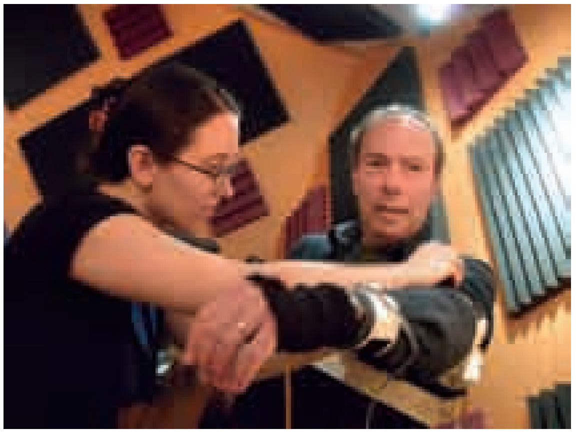
The goal of any live performance, of course, is to lead the audience to concentrate

Elly Jessop and Tod Machover with the robot

on the experience itself, on the ideas and feelings, i.e. to make the "making" of the performance—and in this case all the crazy technology—look simple and inevitable. This is one of the most complex stage shows ever mounted, with numerous individual elements that must work with precision, delicacy, force and beauty, and—even more—function truly as a "system" of interacting machines that must work together in the most unbelievable ways. It is a testament to the opera's production team that for those who don't know, it does look easy. For those of us behind the scenes, it is quite another story of course!

Hopefully, these innovations will lead to new musical possibilities down the line that I can't predict right now, just as software and hardware designed to measure Yo-Yo Ma's bowing led—in a slightly zigzag way—to *Guitar Hero*. I would not be surprised, for example, if the sophisticated infrastructure that Simon Powers uses to construct and communicate his legacy were eventually to morph into a platform for everyone to create and share musical stories—a kind of *Personal Opera*—on your mobile phone, something on which we are already working with several partners, including the Royal Opera House at Covent Garden (London). Just as Simon Powers builds his legacy through

Elly Jessop and James Maddalen



the interconnected elements onstage, *Personal Opera* might provide a new form for preserving and communicating memories, telling stories, and establishing continuity across generations. And we are designing this environment to encourage the spontaneous accumulation of impressions and memories from one's personal databases, shaped by natural input such as vocalizing or gesturing, rather than favoring the *Sargent-Pepper*-like studio production that inspired me so many years ago to find another route. We believe that the result will be both liberating and surprising, and will bring together generations in story telling and listening, and experts and amateurs in a beneficial mentoring environment.

I think that it is precisely this kind of surprising freshness that technology can allow—through what can be precisely customized for each project and through the unexpected new discoveries that each project seems to require or reveal—that remains one of its continuing attractions for me.

But we can't take such freshness for granted. Musical technology is so ever-present in our culture, and we are all so very aware of it, that techno-clichés and techno-banalities are never far away and have become ever more difficult to identify and root out. It is deceptively challenging these days to apply technology to music in ways that explode our imaginations, deepen our personal insights, shake us out of boring routine and accepted belief, and pull us ever closer to one another.

That's what makes this kind of work worthwhile and inspires me. But it also leads to a paradox that I experience every single day: that the desire to shape the future is not perfectly compatible with the knowledge that musical experience—and its power to excite and transform us—is fleeting, here and now, at this very moment. And that we'd be extremely fortunate indeed to create new sounds and instruments and technologies that approach the compact, powerful perfection of playing, listening to or imagining Bach emanating from a solo cello.

What new technology can add to this mix is the potential for establishing a new model for the interrelationship between experts and amateurs in musical listening, performance and creation. Some of the boundaries to active engagement in music have eroded, but there is still much to be done to create a truly vibrant musical culture.

In my view, a prime example of the kind of new musical 'ecology' that we should seek is found in our culture's relationship with cuisine. We all enjoy eating at three-star restaurants and admire the achievements of the world's greatest chefs. At the same time, we do not hesitate to dive in ourselves to prepare special meals of high quality on special occasions. We also put together daily meals for ourselves, improvising content that reflects our personal styles. We enjoy eating and even studying the most 'expert' cuisine we can find, but are not scared to make and invent our own. In turn, the fact that we constantly prepare food ourselves makes us better understand and appreciate other food that we encounter.

Music—and most of the arts—have come very far from such a 'healthy' ecology, and it is this that we need to reinvent. Technology can help, as it can act as a bridge to each of us depending on our background and experience, taking advantage of our skills and compensating for our limitations. Even more importantly, we need to establish a fundamentally new partnership between all of the potential participants in our musical culture, including individual artists, all parts of the music business, technology, lifestyle, health and social organizations, music presenting and broadcasting entities, research institutions, artists-as-mentors and—last but not least—the music-loving public. Only in this way can we establish a culture that will allow music to reach its full potential in shaping and transforming our experience. Doing so will allow music to exert its most powerful possible influence on society at large. Surely we can imagine a world where music—and opera—is at least as nourishing as a three-star meal?

BBVA

Biographies

Curtis R. Carlson, Ph.D. in Geophysical Fluid Dynamics, Rutgers University, is President and CEO of SRI International, Chairman of the Sarnoff Corporation, and honorary Chairman of the Madrid Research Institute in Spain.

He is also a member of America's National Council for Innovation and Entrepreneurship; serves as co-chairman of the Scientific Advisory Board of Singapore's National Research Foundation; is a founding member of the Innovation Leadership Council for the World Economic Forum; and served on President Obama's task force for R&D.

Calrson has received honorary degrees from the Technical University Malaysia, Worcester Polytechnic Institute, Stevens Institute of Technology, and Kettering University. He has served on numerous corporate and governmental boards. He is a member of Sigma Xi and Tau Beta Pi. His many technical publications include *Innovation: The five Disciplines for Creating What Customers Want* (with William Wilmot, Random House, 2006). He holds fundamental patents in the fields of image quality and computer vision and is sought after as a speaker on innovation and global competitiveness.

Alfonso Gambardella (PhD, Stanford, Department of Economics, 1991) is Professor of Corporate Management and Dean of the PhD School of Bocconi University. He is Editor of the *European Management Review* and serves on the Editorial Board of the *Academy of Management Review*, *Global Strategy Journal*, *Industrial and Corporate Change, Research Policy,* and *Strategic Management Journal*. He publishes in several international journals, including *Industrial and Corporate Change*, *Journal of Industrial Economics*, *Management Science*, *Organization Science*, *Research Policy* and *Strategic Management Journal*. He has also published books with leading international publishers, particularly the Cambridge University Press, The MIT Press, and Oxford University Press. His MIT Press book *Markets for Technology* (with Ashish Arora and Andrea Fosfuri) is widely cited. His website is http://www.alfonsogambardella.it/

Hugh Herr is Associate Professor within MIT's Program of Media Arts and Sciences, and The Harvard-MIT Division of Health Sciences and Technology. Professor Herr is the founder and director of the Biomechatronics Resesarch Group at the MIT Media Laboratory. His primary research objective is to apply principles of biomechanics and neural control to guide the designs of wearable robotic systems for human rehabilitation and physical augmentation. In the area of human augmentation, Professor Herr has employed cross bridge models of skeletal muscle to the design and optimization of a new class of human-powered mechanisms that amplify endurance for cyclic anaerobic activities. He has also built elastic shoes that increase metabolic economy for running, and leg exoskeletons for walking load-carrying augmentation. In the area of assistive technology, Professor Herr's group has developed powered orthotic and prosthetic mechanisms for use as assistive interventions in the treatment of leg disabilities caused by amputation, stroke, cerebral palsy, and multiple sclerosis.

Hiroyuki Itami, Ph.D. in Industrial Administration, Carnegie-Mellon University, is Professor of Management and Dean at the Graduate School of Management of Science and Technology, Tokyo University of Science. He has taught (for more than thirty years) at Hitotsubashi University in Tokyo - where he is Professor Emeritus - and also at Stanford Graduate School of Business and INSEAD. He has researched and published widely in various fields including strategy, Japanese corporate systems, and corporate governance, and has recently turned his attention to innovation and a new paradigm of management. He has written more than forty books and many articles, mainly in Japanese but also in English, including *Mobilizing Invisible Assets,* published by Harvard University Press.

Alice Lam, Ph.D., London School of Economics, is Professor of Organization Studies at the School of Management, Royal Holloway University of London. Her research has focused on the relationship between organizational forms, knowledge creation and societal institutions. She also has a long-standing interest in the occupational roles and careers of scientists and engineers. Her current work examines the innovative dynamics of network organizations, and careers and knowledge flows in university-industry collaboration. She has published widely in a range of academic journals including *Organization Studies, Journal of Management Studies*, *Human Resource Management Journal and Industrial Relations*.

Robert E. Litan, Ph.D. Yale University is Vice President for Research and Policy at the Kauffman Foundation in Kansas City and Senior Fellow in the Economic Studies Program at the Brookings Institution. He formerly served in a number of high-level positions including former Director of the Brookings Economic Studies Department; Associate Director, Office of Management and Budget; and Deputy Assistant Attorney General, Antitrust Division, U.S. Department of Justice; as a practicing attorney; and as a consultant on a variety of economic and legal matters to public and private sector institutions. He is the author or co-author of over twenty books and more than two hundred articles on various economic subjects, although primarily on financial institutions and regulation, in professional journals and popular magazines.

Edward Lorenz, Ph.D. in Economics, University of Cambridge, is Professor of Economics at the University of Nice-Sophia Antipolis and a member of the CNRS research Unit, GREDEG. He has held a number of associate and honorary research appointments including Assigned Professor at the University of Aalborg (2003-2008). His research focuses on the comparative analysis of business organisation, employment relations and innovation systems with an emphasis on the European Union member nations. Professor Lorenz has co-edited several volumes including *How Europe's Economies Learn: Coordinating Competing Models* (with Bengt-Åake Lundvall, Oxford University Press, 2006); and *Knowledge, Learning and Routines* (with Nathalie Lazaric, Edward Elgar Press, 2003). He has published numerous journal articles and book chapters including, most recently, "Accounting for Creativity in the European Union: A Multi-Level Analysis of Individual Competence, Labour Market Structure, and Systems of Education and Training", *Cambridge Journal of Economics* (with Bengt-Åake Lundvall, 2010).

Francisco Louçã, Doctorate in Economics and Professor of Economics, ISEG, Technical University of Lisbon. His publications include *As Time Goes By* (with Chris Freeman, Oxford University Press, 2002)-- a book on long waves of development and innovation in modern capitalism, *Turbulence in Economics* (Elgar, 1997) and, recently, *The Years of High Econometrics - A Short History of the Generation that Changed Economics* (Routledge, 2007).

Bengt-Åke Lundvall, M.A. in Economics, University of Gothenburg, is currently Professor of Economics at the Department of Business Studies at Aalborg University and since October 1, 2007 special Invited Professor at Science Po, Paris. In 1992-1995 he was Deputy Director at the Directorate for Science, Technology and Industry, OECD. Lundvall has initiated and co-ordinated networks such as DRUID in Denmark, Cicalics in China and Globelics at the world level. He gives advice on innovation policy in China, France, Denmark, Finland, Sweden, Norway and the Netherlands as well as to international organizations such as UN-WIDER, UNCTAD, UNESCO, UNIDO, World Bank, OECD and in 2009 to the European Commission. The EU nominated him Ambassador for the European Year of Creativity and Innovation. His most recent books are: *How Europe's Economies Learn* (with E. Lorenz, Oxford University Press, 2006), *Handbook on Innovation Systems in Developing Countries* (with C. Chaminade, K. J. Joseph and J. Vang Lauridsen, Elgar, 2009), and *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning* (Anthem, 2010).

Tod Machover has been called "America's Most Wired Composer" by the *Los Angeles Times* and is celebrated for creating music that breaks traditional artistic and cultural boundaries. He is also celebrated for inventing new technologies for music, such as his Hyperinstruments which augment musical expression for everyone, from virtuosi like Yo-Yo Ma and Prince to players of *Guitar Hero*, which grew out of his Lab. Machover is Professor of Music & Media and Director of the Opera of the Future Group at the MIT Media. He has recently launched a Music, Mind and Health group at MIT to develop musical activities that significantly improve a variety of medical conditions and enhance lifelong mental and physical acuity. Machover is particularly noted for his radically inventive operas, the most recent of which – the robotic *Death and the Powers* – premiered in Monaco in fall 2010 under the honorary patronage of Prince Albert II and will tour the US in spring 2011.

Ernesto Martinez-Villalpando is a Ph.D. Candidate at the Biomechatronics Research Group at MIT's Media Laboratory. He holds a Master of Science degree from MIT (2006) and a Bachelor Science in Electrical Engineering from Universidad Panamericana in Mexico (2002). His current area of research in rehabilitation robotics focuses on the development of novel biomimetic lower limb prostheses to improve amputee locomotion. His research objective is to use these biomedical technologies as platforms to understand the principles of human-machine integration. At MIT Ernesto contributed to the development of the world's first powered ankle-foot prosthesis and is currently leading the work to develop a novel active knee prosthesis. Prior to his graduate work in assistive technologies at MIT, he was as a researcher and engineering professor at Universidad Panamericana in Mexico, working in the area of mechatronics and mobile robotics.

Ian Miles is Professor of Technological Innovation and Social Change at the Manchester Institute of Innovation Research, Manchester University Business School, and Co-Director of the Centre for Service Research. Previously he worked at the University of Sussex Science Policy Research Unit.

Much of his work has focused on innovation studies, knowledge-intensive business services and social aspects of information technology. Another closely-related area of activity is Social and Technology Foresight. These studies and other work on research evaluation and social indicators are related to innovation policy-making and policy analysis.

His publications include some twenty authored and edited books and over two hundred journal articles and book chapters.

Recent studies on services include work on services' links to Universities, R&D in services, skill requirements of knowledge-intensive services and services' productivity, as well as a guide to evaluation of innovation programmes, and scenario development in several foresight studies.

Manuel Mira Godinho is Professor of Economics at ISEG, the Economics and Business School of the Technical University of Lisbon. He has taught and carried out research in the areas of Science & Technology Policy, Economics of Innovation, Catching Up and Intellectual Property. He has coordinated several post-graduate programmes in these areas, namely the Globelics Ph.D. School on Innovation Systems and Development. He has authored or been co-author of several books, book chapters and scientific articles published in journals such as *Research Policy*, *Scientometrics*, *Research Evaluation* and *Economics of Innovation and New Technology*. He was awarded a PhD by SPRU, University of Sussex, in 1995 and an MSc by Imperial College, University of London, in 1986.

Frank Moss, B.S. in Aerospace and Mechanical Sciences, Princeton University, M.S. and Ph.D. in Aeronautics and Astronautics, MIT. An entrepreneur and 30-year veteran of the software and computer industries, Frank Moss joined the MIT Media Lab as director in 2006, seeking to make a broader contribution to the world by using technology to address pressing social issues. As head of the Lab's New Media Medicine research group, he focuses on integrating information technologies with health care to radically redefine health-care delivery and empower patients. Moss came to the Media Lab from Infinity Pharmaceuticals, Inc. where he was co-founder and serves on the board of directors. Previously he was CEO and chairman of Tivoli Systems Inc., a pioneer in the distributed systems management field, which he took public in 1995 and subsequently merged with IBM in 1996. Moss also serves on the Board of Trustees of Princeton University. http://www.media.mit.edu/people/fmoss

David Mowery, Ph.D. in Economics, Stanford University, is the William A. and Betty H. Hasler Professor of New Enterprise Development at the University of California, Berkeley Walter A. Haas School of Business, and a Research Associate of the National Bureau of Economic Research. Dr. Mowery taught at Carnegie-Mellon University, was Study Director for the Panel on Technology and Employment of the National Academy of Sciences, served in the Office of the United States Trade Representative and was a Council on Foreign Relations International Affairs Fellow. He has been a member of a number of National Research Council panels, and in 2003-2004 was the Marvin Bower Research Fellow at the Harvard Business School. His research deals with the economics of technological innovation and the effects of public policies on innovation; he has testified before Congressional committees and served as an adviser for the OECD, various federal agencies and industrial firms. Dr. Mowery has published numerous academic papers and has written or edited a number of books, including the *Oxford Handbook of Innovation; Innovation, Path Dependency and Policy; Innovation in Global Industries; Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act;* and *Paths of Innovation: Technological Change in 20th-Century America.* His academic awards include the Raymond Vernon Prize from the Association for Public Policy Analysis and Management, the Economic History Association's Fritz Redlich Prize, the *Business History Review's* Newcomen Prize, and the Cheit Outstanding Teaching Award.

Nashid Nabian, MSc in Architectural Engineering, Shahid Beheshti University; MA in Urban Design, University of Toronto where she received the Toronto Association of Young Architects award; Doctor of Design, Harvard Graduate School of Design. She has been partner at Tehran-based Arsh Design Studio since 2003. Nashid's research focuses on the digital augmentation of architecture and constructed landscapes, particularly public spaces, and on how novel technologies can impact the spatial experience by soliciting the needs and desires of inhabitants or users. She has taught graduate seminar and design studio courses at Toronto University, Rice University and the Massachusetts Institute of Technology. Her research has been showcased in various venues, including the ACADIA, IEEE Digital Ecosystems Conference; the UCMedia Conference on User-Centric Media, the Mobile Multimedia Communications Conference, Toronto's Nuit Blanche Annual Event, New Orleans' American Institute of Architects DesCours Festival, and SEED Magazine. Currently she is a post-doctoral fellow at MIT SENSEable City Lab and a lecturer in the Department of Architecture at Harvard Graduate School of Design.

Alex 'Sandy' Pentland, Ph.D. MIT, directs MIT's Human Dynamics Laboratory and the MIT Media Lab Entrepreneurship Program, and advises the World Economic Forum, Nissan Motor Corporation and a variety of start-up firms. He previously helped to create and direct MIT's Media Laboratory, the Media Lab Asia laboratories at the Indian Institutes of Technology, and Strong Hospital's Center for Future Health. Profiles of Sandy have appeared in many publications, including the *New York Times, Forbes, Harvard Business Review, Newsweek, Caring Magazine (Hospice),* and *Odyssey Magazine (Children).* His most recent book is *Honest Signals,* published by MIT Press. He is among the most-cited computational scientists in the world, and a pioneer in computational social science, organizational engineering, mobile computing, image understanding, and modern biometrics. His research has featured in *Nature, Science,* the *World Economic Forum, Harvard Business Review, Newsweek,* the *New York Times,* Vogue, O Magazine, and the National Inquirer, as well as being the focus of dozens of TV features including *Nova* and *Scientific American Frontiers.*

Carlo Ratti, an architect and engineer by training, practices in Italy and teaches at the Massachusetts Institute of Technology, where he directs the SENSEable City Lab. He graduated from the Politecnico di Torino and the École Nationale des Ponts et Chaussées in Paris, and later earned his MPhil and PhD at the University of Cambridge, UK. Ratti has co-authored over 100 scientific papers and holds several patents. His work has been exhibited worldwide at venues such as the Venice Biennale, the Design Museum Barcelona, the Science Museum in London, GAFTA in San Francisco and The Museum of Modern Art in New York. His Digital Water Pavilion at the 2008 World Expo was hailed by Time Magazine as one of the Best Inventions of the Year. He has been included in

Esquire Magazine's Best and Brightest list and in Blueprint Magazine's 25 People who will Change the World of Design. Ratti recently served as the inaugural Innovator in Residence in Queensland, Australia.

Nathan Rosenberg, Ph.D. Wisconsin, honorary doctoral degrees, Universities of Lund and Bologna, is the Fairleigh S. Dickinson Jr., Professor (Emeritus) of Public Policy in the Department of Economics at Stanford University. He was educated at Rutgers University, University of Wisconsin and Oxford University. His books include *The American System of Manufactures, Perspectives on Technology, Inside the Black Box, Technology and the Pursuit of Economic Growth* (with David Mowery), *How the West Grew Rich* (with L. E. Birdzell, Jr.), *Exploring the Black Box*, *The Emergence of Economic Ideas*, *Schumpeter and the Endogeneity of Technology: Some American Perspectives, Paths of Innovation: Technological Change in 20th-Century America* (with David Mowery).

Nathan Rosenberg has served as chairman of the Stanford Economic Department. He is a member of the Board of Directors of the National Bureau of Economic Research, chairman of the advisory board of the UN Institute for New Technology, and a fellow of the Canadian Institute for Advanced Research. He is an Elected Fellow of the American Academy of Arts and Sciences and the Swedish Royal Academy of Engineering Sciences.

Edward S. Rubin, Ph.D. Stanford, is a professor in the Departments of Engineering & Public Policy and of Mechanical Engineering at Carnegie Mellon University. He holds a chair as The Alumni Professor of Environmental Engineering and Science, and was founding director of the university's Center for Energy and Environmental Studies and the Environmental Institute. He is a Fellow Member of the ASME, recipient of the AWMA Lyman A. Ripperton Award for outstanding achievements as an educator, and the Distinguished Professor of Engineering Award from Carnegie Mellon. He serves on advisory committees to government agencies and was a coordinating lead author of the 2005 Special Report on Carbon Dioxide Capture and Storage by the Intergovernmental Panel on Climate Change (IPCC), co-recipient of the 2007 Nobel Peace Prize.

Takanori Shibata, Ph.D. in Electronic and Mechanical Engineering, Nagoya University, is a Senior Research Scientist at the National Institute of Advanced Industrial Science and Technology in Japan, having been a research scientist at the Artificial Intelligence Lab., MIT, and a visiting research scientist at the Artificial Intelligence Laboratory, University of Zurich. He was Deputy Director for Information and Communication Technology Policy at the Japanese Government Bureau of Science, Technology, and Innovation Policy from 2009 to 2010. His research interests include human-robot interaction, robot therapy and humanitarian de-mining. He was certified by the Guinness World Records in 2002 as the inventor of a seal robot named Paro, the World's Most Therapeutic Robot. His awards include the Robot of the Year award from the Japanese Ministry of Economy, Trade and Industry, the Outstanding Young Person in the World award from the Junior Chamber International, and the Japanese Prime Minister's Award.

Pascal Soboll, M.A. in Engineering (Product Design), Stanford. As a practice lead at IDEO's Munich office, Pascal works with clients to find and tap into opportunities for growth. He helps major enterprises, among them P&G, Daimler and BBVA to shape strategies and then guide them through the concept stage and all the way to implementation. He is particularly interested in the areas of energy and mobility, where technology, human needs and sustainability merge to open up new ways forward. Pascal worked for IDEO in the U.S. and the UK before returning to his native Germany. Previous jobs include scientific research for Daimler and Stanford University and car design for GM/Opel. He holds a diploma in Theoretical Physics.

A self-confessed car nut, his collection of project cars has recently shrunk in inverse proportion to his growing family.

Sander E. van der Leeuw, Ph.D., University of Amsterdam, an archaeologist and historian by training, has taught at Leyden, Amsterdam, Cambridge (UK) and Paris.

Van der Leeuw's core research theme is the study of socio-environmental dynamics, but he has also contributed to the reconstruction of ancient technologies, (ancient and modern) regional man-land relationships, GIS and modeling. Since 1981, he has worked on applying Complex Systems Theory in various ways in all these domains. From 1992 to 1999, he coordinated a series of major research projects funded by the EU on (modern) socio-environmental problems in Southern Europe. He is currently involved in applying Complex Systems approaches to the study of invention and innovation in the United States and Europe, funded by the Marion Ewing Kauffman Foundation and DG ICT of the European Commission

Since late 2003, he has been Professor of Anthropology at Arizona State University, and Director of the School of Human Evolution and Social Change. As of July 1, 2010, he simultaneously fills the position of Dean of ASU's School of Sustainability.

Joaquim Vilá, Ph.D. in Strategic Management, Wharton School, University of Pennsylvania, is Professor of Strategic Management and Innovation, and Academic Director of Executive Programs on Innovation Management at IESE Business School. Consultant in the implementation of systematic approaches to foster innovation in companies from different industries, both industrial and services. Member of Advisory Boards and Councils Innovation in pioneering companies. Co-founder of three companies that offer services in various areas to promote innovation (Soft4CRIT-CRITflow, Estratègia i Organització, and Total Business Innovation). Regular speaker on innovation in management training programs, in-company programs for advanced corporations and international forums. His publications focus on management issues such as how to promote broad, robust and systematic innovation (Business Innovation), the development of innovative competences in management teams, and how to make innovative strategies relevant to middle management to encourage a successful implementation.

Xavier Vives, PhD in Economics, University of California, Berkeley, is a professor of economics and finance, holder of the Abertis Chair and academic director of the Public-Private Research Centre at IESE Business School. He has taught at the Institut d'Anàlisi Económica (Spanish National Research Council-CSIC), the Institució Catalana de Reserca i

Estudis Avançats (Universitat Pompeu Fabra), INSEAD, Harvard University, New York University and the University of Pennsylvania. He is a fellow of the Econometric Society and the European Economic Association, and a member of the Economic Advisory Group on Competition Policy of the European Commission and the European Economic Advisory Group at CESifo. He has also worked as director of the Industrial Organization Programme at the Centre for Economic Policy Research and editor of the *Journal of the European Economic Association*. He has published numerous articles in international journals and is the author of several books, the most recent being *Information and Learning in Markets.* He has received a number of distinctions in the course of his career, including the European Research Council Advanced Grant (2008).

Eric von Hippel, Ph.D. in Innovation, Carnegie-Mellon University, is T. Wilson Professor of Innovation Management and also Professor of Engineering Systems at MIT. He is known for his research into the sources of innovation. He finds that product development is rapidly shifting away from product manufacturers to product "lead users" in the Internet Age. The rapid growth of innovation by lead users requires major changes in company business models and government policymaking. Von Hippel's new book, *Democratizing Innovation* (2005), explains user-centered innovation and how companies and nations can adapt and profit. This book is available free on the web at http://mit.edu/evhippel/www/books.htm

Harry West, Ph.D. in Mechanical Engineering, MIT, guides the strategic direction and global growth of Continuum. He is an experienced innovation practitioner: engaging with clients, understanding global consumers, and designing innovation. He travels and learns constantly to stay connected with real needs in our rapidly changing world.

Harry West has worked with a wide variety of brands in a broad range of industries, including American Express, Andersen Windows, AstraZeneca, BBVA, BMW, Fidelity, Master Lock, Procter & Gamble, and Sprint. For P&G he led teams that helped to create Swiffer® and ThermaCare®, and the new design language for Pampers diapers. Under his leadership, his teams have received IDEA, ID and Red Dot awards, and numerous design and utility patents.

Harry West is a 16-year veteran of Continuum and was named chief executive officer in 2009. Prior to joining Continuum, Harry was associate professor of mechanical engineering at The Massachusetts Institute of Technology (MIT) where he taught design and control.

We would like to thank the following people for their kind collaboration: Coloma Casaus; Eva Mendo; Galerie Chantal Crousel, Lara Blanchy; Galerie Urs Meile, Karin Seiz; Gladstone Gallery, Sascha Crasnow; Hauser & Wirth, Maria de Lamerens; IVAM Institut Valencià d'Art Modern, María Casanova, Mercedes Lerma; Kapoor Studio, Clare Chapman; Luise Kaunert; Joyce Lai, Marga Paz; Marian Goodman Gallery, Catherine Belloy, Carole Billy; MIT Museum, Claire Calcagno; Taller Fontcuberta, Mar Sorribas Roca; The Institute of Contemporary Art, Boston, Colette Randall; and Wingate Studio, Peter Pettengill.

Publisher
BBVA

Project direction and coordination
Chairman's Advisory, BBVA
Communication and Brand, BBVA

Texts
Curtis R. Carlson, Alfonso Gambardella, Hugh Herr, Hiroyuki Itami, Alice Lam, Robert E. Litan, Edward Lorenz, Francisco Louçã, Bengt-Åke Lundvall, Tod Machover, Ernesto Martínez-Villalpando, Ian Miles, Manuel Mira Godinho, Frank Moss, David Mowery, Nashid Nabian, Alex 'Sandy' Pentland, Carlo Ratti, Nathan Rosenberg, Edward S. Rubin, Takanori Shibata, Pascal Soboll, Sander E. van der Leeuw, Joaquim Vilá, Xavier Vives, Eric von Hippel and Harry West.

Publishing coordination
Tf Editores

Graphic design
Juan Antonio Moreno. Tf Media

Translations
Polisemia

Text editing
Ann Uldall

Research
Mercedes Abstengo

Pre-printing
Cromotex

Printing
Tf Artes Gráficas

Binding
Ramos

Cover: *Eden* (2004-2008) by Jon McCormack

D.L.: M-51581-2010
Printed in Spain