Essays on Social Learning and Imitation

by Bulat Sanditov

Contents

Chapter 1

Introduction

Paying in a restaurant the Dutch often split the bill according to their orders. In the same situation, Italians may choose to go for *pagare alla romana* paying in equal amounts.[1] In Balkans people show their approval shaking their heads; a nod is considered as a sign of disagreement. In many other places it is exactly the opposite. The Japanese and the British ride on the left side of a road. The rest of the world rides on the other side. These and numerous other examples of "cultural differences" illustrate that, on the one hand, there is a multitude of ways to address similar situations which are embedded in everyday routines, traditions, and institutional settings. On the other hand, there is a striking conformity of behavioural patters within each of the groups (neighbourhood, nation, social class).

The mechanisms leading to conformity are many. In this thesis I deal with cognitive (social learning) and normative (social norms) factors resulting in performing a wide spectrum of tasks "as others do". This thesis explores some economic contexts in which these factors play a crucial role. The plan of the dissertation is as follows. In **Chapter 2** we start with a review of literature on social learning and imitation, our aim there is to give some broad overview of the body of research done in a range of behavioural sciences on these issues, as well as provide a background for four particular economic contexts examined in the rest of the thesis.

Chapter 3 enquires in the relationship between the structure of information flows and efficiency of social learning. To address this problem we develop and analyse a model of social learning in networks following the lines of Bala and Goyal (1998,2001). They have examined a model of social experimentation, where agents can observe actions taken by their closest neighbours and their outcomes, and found that the higher is the "degree of integration" within the society is (density of the communication network), the more likely it is that the optimal action (with respect

[1] A translation verbatim from Italian *pagare alla romana* is "to pay as Romans do". Somewhat counterintuitvely, in Rome *pagare alla romana* means to pay for oneself. In the rest of Italy it means to split the bill equally.

to payoff) prevails, and, as a result, conformity of actions will arise. However, we argue that when the information is not shared (i.e. only actions are observable), and decision is not reversible the conclusions may be different.

Chapter 4 is inspired by spectacular rise and burst of "dot-com" bubble in the last decade of the 20th century. Interestingly, both sides of the market: the entrepreneurs founding "dot-com" start-ups and investors pouring money into "dot-com" ventures, happened to be overoptimistic about the future of the "new economy". In this chapter we argue that social learning in an environment with asymmetrically distributed information may lead to "mutual illusions" on the both sides of the market.

Chapter 5 presents a simple evolutionary model to study the diffusion patterns of product innovations for consumer goods. Following a Veblenian theme we interpret consumption as a social activity constrained by social norms and class structure. We apply Veblenian analysis to explaining the diffusion of a new product, and formulate a model based on replicator dynamics to study the effects of class structure and social norms on characteristics of the diffusion process (diffusion speed and market penetration). We frame our argument as a contribution to the debate on the role played by demand in the process of industrialization in the Western Europe.

Chapter 6 touches upon the issues of path-dependency in the process of technical change. In the evolutionary theory the path-dependency is often attributed to the existence of "focusing devices / technological guideposts / technological trajectories", i.e. the theory emphasises the role of the cognitive factors framing the direction of innovation search. From our perspective we can easily recognize the elements of the social learning behind those factors. In this chapter we examine how the framework of path-dependency and technological trajectories can be applied to explain the observed distribution of patent values as it is revealed in the distribution of patent citations.

The last chapter concludes the dissertation.

Chapter 2

Review of Literature on Imitation

2.1 Introduction

In the preface to his volume on *Extraordinary Popular Delusions And The Madness Of Crowds* published in 1841 Charles MacKay wrote, "In reading history of nations, we find that, ... whole communities suddenly fix their minds upon one object, and go mad in its pursuit; that millions of people become simultaneously impressed with one delusion, and run after it, till their attention is caught by some new folly more captivating then the first. ... Men, it has been said, think in herds; it will be seen that they go mad in herds, while they only recover their senses slowly and one by one" (MacKay 1841). These words are as true today as they were 165 years ago. To say more, as in the course of the last centuries the world has become more connected, and frenzies, fads and fashions are getting global, perhaps these words are even more relevant today than ever before.

The major reason the history of the mankind is so abundant with the stories about "madness of crowds" so well documented by Mackay is rooted in the social character of our being. Social interactions shape our way of thinking, expectations, judgements and, ultimately, our behaviour. It is so natural that most of the time it goes unnoticed, but under certain conditions the social component reveals itself so openly that it cannot be ignored. The particular feature that allows frenzies to unfold is human readiness and ability to copy behaviour of others. Social nature and constitutional bias towards imitation govern crowds.

The importance of imitation, however, goes far beyond the phenomenon of fads and fashions; some might argue that the very survival of our species is due to our special gift for copying the behaviour of others. Because of its importance, imitation has been studied intensively. The literature on imitation is as vast as it is diverse – it spreads throughout a range of disciplines: biology, anthropology, psychology, sociology, and, more recently, economics.

There are good reasons for economists to study imitation. There is a variety of economic contexts where imitation plays an essential role. To mention only a few: herding in financial markets is a significant component of investment bubble dynamics, it contributes to both growth and consequent burst of the bubble; imitative strategies employed by firms have important implications for industrial dynamics; with respect to R&D, imitation alongside innovation is the driver of technical change and long-run economic growth.

This chapter provides a short review of the literature on imitation. The next section attempts to outline some of the directions of imitation research in several behavioural sciences. Section 2.3 is a review of studies on imitation within current mainstream economic theory. Section 2.4 highlights several themes on social learning and imitation that will be explored in detail in this dissertation.

2.2 Imitation in behavioural sciences

The purpose of this section is to put the discussion of economic aspects of imitation and its consequences in the broader context of behavioural sciences. We start with two instances of imitation documented by students of animal behaviour. These stories illustrate the fact that imitation is not specific to humans but is common in the animal world, as imitation is a tool for adaptation to the environment and, ultimately, for survival in the course of natural selection. Therefore to some degree our natural inclinations and abilities to imitate are 'hard-wired' in us, if not genetically then culturally. Sociobiology and memetics theories are mentioned in this context. It is important to note that imitation occurs not only on the level of individuals; imitation is also an important element of organization behaviour. Organization theory will be discussed more extensively because at the moment the most systematic analysis of economic aspects of imitation has been done within this field.

2.2.1 Individual level

Biology In 1981 Ran Aisner, a high school biology teacher, was on a field trip with his class in the pine forests in northern Israel when he observed a pile of cone shafts beneath certain pine trees. The pattern in which the pine cones had been stripped of their scales was characteristic of a squirrel; however no squirrel has been ever reported to be seen in Israel forests!

A subsequent study by Aisner and Terkel revealed that the trees in these forests are inhabited by black rats (*Rattus rattus*), which feed on the pine seeds (Aisner and Terkel 1992, Terkel 1996). Pine seeds are the main source of their nourishment in the relatively sterile pine forests of northern Israel. Rats are known as opportunistic feeders; however to be able to feed on pine seeds an animal needs to know how to strip

a cone in an efficient way, otherwise stripping, chewing, and digesting takes more energy than the nourishment value of the cone's seeds. How do rats evolve to strip pine cones, if the mechanism of natural selection usually operates on a different time scale? Pine forests of northern Israel lack typical species characteristic of similar habitants in other parts of the world — there are no old natural forests on the territory of modern Israel, at the time the oldest ones were 40 years old — therefore Aisner's finding came as much of surprise.

Laboratory experiments by Terkel and collaborators have shown that the skills are neither inherited, nor learned through trial-and-error, but transmitted between generations from dam to pups through observing dam's behaviour. In other words, one can speak of cultural rather than genetic transmission(Terkel 1996).

An early account of social learning, often mentioned in the literature on animal culture, describes spread of bottle openiing in birds. In the 1920s small birds called titmice or tits, common in British gardens, were seen prising open wax-sealed tops of milk bottles left on doorsteps. Although the distribution of the milk by milkmen delivering bottles to the doors of each country house had been practiced for a long time, titmice pecking the top of a bottle in search of cream had never been seen before. The pattern emerged locally in the south of England, and later spread gradually from one place to another across England and then to some parts of Scotland and Wales. Even when later milk producers started to close the bottles with aluminium foil birds learned to open foil tops as well (Fisher and Hinde 1949).

These and other instances thoroughly documented by researchers working in the field of animal behaviour (Heyes and Galef Jr. 1996) highlight the essential role that *social learning and imitation* play in propagating behaviour that allow animals to occupy an ecological niche which might otherwise be closed to them (as in the case of rats in forests of northern Israel).

Of course, imitation as an effective mechanism for transmission of acquired behaviour, skills, and tradition serves an important function not only in the animal world, but also in human societies. And it has not gone unnoticed. The ability of humans to imitate others was stressed as early as Aristotle: "Imitation is natural to man from childhood, one his advantages over the lower animals being this, that he is the most imitative creature in the world, and learns at first by imitation."(cited in Meltzoff (1988))

Social learning theory Although learning from observing the behaviour of others had been documented by many authors (consider Aristotle, or Gabriel Tarde's "laws of imitation"), a comprehensive theory of social learning was first proposed by Albert Bandura (1977), a professor of psychology at Stanford university. Earlier theories of human behaviour focused on the process of learning that takes place within the individual. In those theories the process of learning, i.e. acquiring a

novel mode of behaviour, consists of series of trial-and-errors where different actions of the individual are rewarded or punished by the environment. Individuals learn how characteristics of the environment are related to the rewards of their actions. Hence environment causes individual behaviour.

The social learning theory extends the scope of analysis: it looks outside of the individual, and considers interaction between behaviour and environment. Environment via learning causes individual behaviour, but the individual behaviour, in turn, causes environment ('reciprocal determinism'). Causality flows in both directions: from environment to individual, and from individual to environment.

As with other theories of human behaviour, social learning theory acknowledges that some learning occurs through direct experience and consequences of individuals actions that reinforce certain behavioural patterns. However, it also considers another way of learning — through observing actions of others, and conjectures that complex patterns of social behaviour are more likely to be acquired through this channel:

> Although behavior can be shaped into new patterns to some extent by rewarding and punishing consequences, learning would be exceedingly laborious and hazardous if it proceeded solely on this basis ...it is difficult to imagine a socialization process in which language, mores, vocational activities, familial customs, and the educational, religious, and political practices are taught to each new member by selective reinforcement of fortuitous behaviors, without benefit of models who exemplify the cultural patterns in their own behavior. Most of behaviors that people display are learned either deliberately or inadvertently, through the influence of example. (Bandura (1976, p.6) cited in Davis and Luthans (1980))

Observational modelling consists of several steps: attention (to learn the individual has to pay attention to the model's behaviour), retention (the individual stores information as a mental image or verbal description of the situation), reproduction (the individual should be able to reproduce model's behaviour), and motivation.

Observational learning does not require any verbal exchange of information between the individual and the model — the individual can extract information just from observing a model's behaviour. Therefore observational learning might save not only on the cost of redundant experimentation, but also on costs of direct communications (costs of establishing and maintaining personal contact, codifying the model's knowledge etc.) In this way observational modelling may happen via mass-media such as newspapers, TV, or Internet.[1]

Researchers working in the fields of animal behaviour and social psychology distinguish between the two terms: social learning and pure imitation. Broadly defined

[1]Laboratory experiments of Bandura and his students, in particular, the famous "bobo doll" experiment, raised debates about potential danger of displaying violence on TV.

social learning is a process of learning something from observing the behaviour of conspecifics; *imitation* is usually considered as a particular case of social learning. According to Heyes (1993) pure imitation occurs when imitators learn about the form of behaviour from observing others, while other forms of social learning are learning about the environment from observing others. Blackmore (1998) defines imitation in terms of copying: imitation takes place only when some (novel) form of behaviour is copied from model to imitator.

There are other modes of social learning besides imitation. For example, the spread of milk-bottle pecking among tits mentioned above is not true imitation. Indeed, observing other birds a tit learns that it is worth-while to try pecking at the top of a milk-bottle. When it manages to peck through the top, it learns that there is cream under it. This process of social learning is not imitation, because the bird does not learn how to peck, pecking is natural to tits, and indeed observers reported that birds used different methods to open the bottles.[2]

SocioBiology and Memetics The role of true imitation as an efficient mechanism in cultural transmission is emphasised in sociobiology, a branch of evolutionary biology studying the process of coevolution of genes and social structure (culture), and in particular in the theory of meme, that is a theory of evolution of human societies based on a neo-Darwinist approach. The latter considers any particular culture based on a set of "memes" that are "stories, songs, habits, skills, inventions, and ways of doing things that we take from person to person by imitation" (Blackmore 2000).[3]

Although imitation is not the only mechanism for social transmission of novel behaviour it does have special features that other modes of social learning do not have. Blackmore (1998) states:

> Although new behaviours can be passed on by other kinds of social learning, the process is cumbersome. For example, one animal must invent a new behaviour during individual learning and then somehow lead a second animal into such a situation that it is likely to learn the same new behaviour - or perhaps the first can behave in such a way as to change the contingencies of learning for the second animal so that it learns the same (or a similar) new behaviour. Most importantly, in these cases, the behaviour must be created anew each time by the learner. The social situation, and the behaviour of the other animal plays a role, but the details of the first behaviour are not transmitted and therefore cannot be built upon and refined by further selective copying. In this sense, then, there is no true heredity.

[2]This form of social learning is called "stimulus enhancement", because in this case birds learn to pay attention to stimulus, the bottle top.

[3]Critics of this approach notice that "meme" or similar concept of "culturgene" are "merely neologisms for that time-worn unit of classic diffusionist ethnology, the culture-trait." (Ingold 1990)

By contrast, imitation is an efficient way for new models of behaviour (memes) to be transmitted from one individual to another rapidly and without much loss of information. From the point of view of memetics, the ability of a meme to reproduce itself via imitation makes it another "replicator" (Dawkins 1976), similar to genes. Therefore, memetics sees human evolution as a co-evolutionary process with two replicators: a gene that defines our physical shape, and a meme that determines our behavioural models. Most sociobiological theories neglect the latter. They assume that it is the genetic evolution that defines our species culture. O.E. Wilson famously said that the genes hold culture on a leash (cited in Blackmore (2000)). By contrast, memetics emphasises that the interaction between the two replicators runs both ways, and unfolding cultural evolution encoded in memes is important for determining which genes are to be selected. For example, Blackmore (2000) conjectures that this co-evolutionary process may explain the excessive brain size[4] characteristic to humans:

> It is easy to imagine that our early ancestors imitated useful new skills in making fire, hunting, and carrying and preparing food. As these early memes spread, the ability to acquire them became increasingly important for survival. In short, people who were better at imitation thrived, and the genes that gave them the bigger brains required for it consequently spread in the gene pool. Everyone got better at imitation, intensifying the pressure to enlarge the brain still further in a kind of cerebral arms race. (Blackmore 2000)

And in this way humans have become "fundamentally unique not because they are especially clever, not just because they have big brains or language, but because they are capable of extensive and generalised imitation" (Blackmore 1998).

2.2.2 Organization behaviour

One of the conclusions we draw from the literature is that ability to imitate is an important factor for species survival, as it provides an efficient channel through which experience can be transmitted from one individual to another. The pressure on natural selection guarantees that only beneficial behavioural traits essential for survival will dominate in the long run. However the selection process works not only on the level of individuals. Selection also operates on the level of groups of individuals, such as colonies, families, or, in an economic context, firms.[5] This kind of perspective has been taken up by organization theory.

Organization theory recounts that a firm is "an *adaptively rational* system rather than an *omnisciently rational* system" (Cyert and March 1963). It assumes that

[4] "Excessive" with respect to the "minimum requirements for survival".

[5] In evolutionary game theory this can be formalized as "group selection" games. For example, see the 'haystack' model by Smith (1964).

a firm's operations follow some standard operational procedures, in that a member of an economic organization (and, in fact, any organization) performs her functions according to "task performance rules" (Cyert and March 1963) or "routines" (Nelson and Winter 1982). These rules are "behavioural traits" of organizations and their role in the process of economic selection is similar to the role genes play in natural selection. The importance of imitation stems from its function in the mechanism of transmission and multiplication of routines within and between organizations.

Intra-organizational learning Organization theory asserts that task-performing rules have an evolutionary origin; they arise in the process of boundedly rational search: through a process of trial-and-error members of an organization develop some heuristics to deal with particular kinds of problems they face in their position in that organization. However, would the process of organizational learning stop there, it would be rather unfortunate for at least two reasons.

First, since the rules are derived from experience rather then from "first principles" (e.g. profit maximization) without institutionalized transmission of rules within units of the organization, any rotation of personnel would result in partial or complete loss of organizational memory. It would imply that as personnel changes, everyday routines would have to be reinvented over and over again. As Cyert and March (1963) put it:

> Consider a new employee in an organization who is given the simple instruction, "Set price so as to maximize profit." If such an employee lasted long enough in the organization — and the organization lasted long enough — some ways for handling the pricing problem that were reasonably satisfactory would eventually be developed. But presumably prior employees have dealt with the same problem and developed some procedures. The organization's rules permit the transfer of past learning. (Cyert and March 1963, p.104)

A new employee does not have to solve some formal optimization problem or go through the process of trial-and-error to figure out how to perform his functions, instead relevant rules can be learned from observing the work of her colleagues.

Second, a routine developed in a particular organizational unit as a response to a specific task might happen to have more general applicability than the task itself, and if shared with other units, it could potentially be beneficial to the organization as a whole. For example, this might be the case when an organization has to respond to systemic changes in the environment, and the same kind of changes have to be dealt with by different units. Efficient institutionalised transmission of ideas between units of an organization would ensure that some general procedure will be developed and shared; otherwise valuable resources will be wasted on duplicate efforts.

The importance of sharing of experience within organizations is well recognized in practice

> Even if you have developed a high-caliber system of innovation, you will *still* not have institutional learning until you develop the ability to "flock". [...] Some managers see conventional training and development as merely an opportunity to acquire some new skills. However, ... training and development becomes a powerful vehicle for institutionalizing learning. [...] The flocking is intensive; course attendees nearly always tell you afterwards, "It is not so much what I learned in the official sessions, but what I picked up from my colleagues during the breaks that was important. (de Geus 1997)

Moreover, interaction between different units is important for learning to spread through the organization, otherwise

> [w]e should therefore not be surprised, when these teams communicate antagonistically ..., at squabbling at the boundaries of their territories. The amount of institutional learning is limited ... the chances that the innovative ideas will become company policy are much reduced." (de Geus 1997)

Information exchange and learning in organization may take different forms. An employee may learn her functions and the way to carry them out via formal contacts such as sets of instructions, formal communications with her managers and so on, but as the quote above suggests the most important information comes via informal channels, from talking to and observing other employees. Social learning is particularly important in this mode of information transmission. According to the social learning approach to organizational behaviour, observational modelling plays an important role in transmitting task performance rules (as well as other formal and informal practices) among organization members:

> ... organizational participants learn how to behave from observing those around them. The dictum "Do as I say, not as I do" seems unlikely to be followed. Job description, rules, and policies are more likely to be interpreted from watching what others do than following written directives. The example by behavior that managers provide for their people may be more important than instructions they provide (Davis and Luthans 1980)

Thus social learning and imitation are important mechanisms through which organizations accumulate experience, and share and sustain task-performance rules.

Inter-organizational learning Besides being an important mechanism for transmission of behaviour *within* an organization and therefore affecting the efficiency of organizational learning, social learning also plays essential role in the process of *inter-organization* learning. First, as in the case of intra-organizational learning, the ability to copy behaviour of the others helps by saving on the costly process of search. In addition, "standardization" of task performing rules on the level of population of organizations permits managers to "read" behaviour of other firms, and to avoid uncertainty:

> [S]ome rules are more general than the individual firm and are identified as a more pervasive code called "industry standard practice", "standard business practice", "ethical business practice" or "good business practice." ... "good practice" – especially at managerial levels – tends to be shared among firms. ... [general rules] serve the important function of providing an operational procedure for the manager to use in a situation of comparative ambiguity. Insofar as the external situation consists in other firms, it becomes predictable. Competitors' behavior can be predicted in the areas covered by standard practice. Insofar as potential failure is of concern to the decision maker, "standardization" provides defense. (Cyert and March 1963, p.105)

Adoption of a "good business practice", whether it is related to new production or management technology, is an instance of imitation.

Organization theorists distinguish between three fundamental modes of imitation: frequency-based imitation, trait-based imitation, and outcome-based imitation (Haunschild and Miner 1997). *Frequency-based imitation* occurs when organizations try to implement practices widely adopted in the industry. Fligstein (1985) examined the spread of the multidivisional form (MDF) of organization among large firms and found that the probability of adoption is increasing with the share of other firms in the same industry that adopted MDF. Palmer, Jennings, and Zhou (1993) focused on the adoption of MDF in a sample of late adopters (with a different set of other explanatory variables) and came to the same conclusions with respect to the effect of the prevalence of the MDF in the respective industry. Burns and Wholey (1993) analyzed adoption of matrix management by hospitals and found that the probability of adoption increases with the cumulative share of adopters among local hospitals. Rao, Greve, and Davis (2001) studied initiation of coverage of firms listed on the NASDAQ by investment analysts and found that the number of analysts initiating coverage of a certain stock in the previous year positively affects the probability of this stock to attract another analyst.

In contrast to frequency-based imitation, *trait-based imitation* is focused on practices in use only by a subset of other organizations having specific characteristics associated with high status. These characteristics usually include size, prestige, and success. Results of Burns and Wholey (1993) suggest that diffusion of matrix management proceeds from higher- to lower- prestige hospitals. Haveman (1993) studied the Californian industry of savings and loan and found positive relationship between the decision to enter into a new market and the presence of large and profitable firms in this market.[6] In another study Rao, Greve, and Davis (2001) showed that decision to start coverage of a stock depends on the number of high-status analysts started to cover this stock in the previous year. When making the choice

[6]The relationship is positive if the share of high-status firms does not exceed a certain limit. Above this limit, the relationship turns negative due to the effects of competition.

of a model for imitation, an organization might take into account not only the traits associated with success, but also the degree of similarity between the model and the organization. Baum, Li, and Usher (2000) studied acquisitions by chain nursery homes in Canada with respect to geographical location of acquisition targets. In the line with the other studies mentioned above they found that choices of larger chains are likely to be imitated immediately. In addition, they found that choices of similar-sized chain organizations are also likely to be imitated, but with some lag and conditional on the success ('wait-and-see' strategy.).

With *outcome-based imitation*, organizations copy practices that brought favourable outcomes to organizations which use them, and avoid practices that led to bad outcomes. Most diffusion studies confirm that diffusion (imitation) rate depends on profitability (Griliches 1957, Mansfield 1961). Therefore if we assume that the knowledge about profitability of technology (practice) spills over to other firms, then this can be considered as a form of outcome-based imitation. There are also empirical studies that indicate that salient positive outcomes are more likely to be imitated. For example, Conell and Cohn (1995) found that strikes in the coal mining sector in France did stimulate other strikes in the same department, and the impact was significantly higher if the initial strikes were victorious. Haunschild and Miner (1997) examined the factors that influence companies' choices of an investment bank as an adviser on an acquisition and found that highly salient outcomes sustain outcome-based imitation mode.

The literature suggests different theoretical rationales for different modes of imitation. DiMaggio and Powell (1983) argue that there are three main mechanisms of isomorphic change (the process leading toward homogenization). First, there is *coercive isomorphism*, that results from formal and informal pressures exerted on an organization by other organizations on which the organization is dependent (e.g. government, financial institutions, suppliers). Adoption of "greener" technologies in order to comply with environmental standards enforced by the government is the case in point. Another mechanism is *normative pressure* that stems primarily from professionalization. For instance, Palmer, Jennings, and Zhou (1993) studied adoption of MDF by large US corporations in 1960s and found that having a CEO with an MBA from an elite business school increased the probability of adoption of MDF. Last but not the least, there is *mimetic isomorphism* that arises as a response to uncertainty surrounding a novel practice/technology. In such an environment organizations may model themselves on other organizations. This mechanism of isomorphic change is directly related to trait-based imitation modes:

> Organizations tend to model themselves after similar organizations in their field that they perceive to be more legitimate or successful. The ubiquity of certain kinds of structural arrangements can more likely be credited to the universality of mimetic process than to any concrete evidence that the adopted models enhance efficiency." (DiMaggio and Powell 1983, p.152)

Similarly, as the passage from Cyert and March (1963) quoted above suggests, the legitimacy of a "good business practice" may stem from its widespread use, and therefore an organization may choose to imitate the practice prevailing in the field, i.e. in this case we speak of the frequency-based imitation mode.

While frequency- and trait- based imitation modes are due to institutionalised pressure toward homogeneity, outcome-based imitation is related to the economic value of the practice/technology. Organization theorists have argued that just as individuals can learn from observing actions of others (social learning theory), organizations are also able to learn from the experience of other organizations and make decisions to imitate or to avoid practices depending on their perceived impact on earlier adopters (Davis and Luthans 1980). This view on imitation is in a way similar to the standard treatment of adoption process in the economics literature, and it will be discussed in detail in the next section.

2.3 Imitation in Economics

Historically, mainstream economic theory with its focus on formal market mechanisms of exchange and emphasis on perfectly rational individual agents largely ignored non-market mechanisms. However, the last two decades have seen growing recognition of the importance that non-price interactions play in economic activities. As a result there is an increasing number of attempts to fit social interactions in general and imitation mechanisms in particular into the world of formal economic models.

2.3.1 *Homo economicus*

The economic model of human behaviour known as *homo economicus* is very distinct from the models of individual behaviour developed in other behavioural sciences. She is an amazing creature in all ways. Her *raison d'etre* is to maximize her utility at all times. For that, in addition to a comprehensive and coherent set of preferences, she possesses brain power exceeding all imaginable limits, and never hesitates to apply it in any circumstances. Organizations consisting of many *homines economici* together share most features with the individual *homo economicus*: omniscient rationality and greediness.

What is a reason for perfectly rational individuals or organizations to copy behaviour of other individuals? While other behavioural models reviewed in the previous section consider the tendency to imitate as "hardwired" in boundly rational individuals and organizations in the course of evolution, this cannot be the case with perfectly rational *homo economicus*: she must be sure that she is better off with copied behaviour than any other alternative. Therefore, observed imitative be-

haviour should be explained in terms of individual (expected) payoff from copying behaviour of the others.

Before we start reviewing economic models of imitation, we shall note that economists are cautious not to interpret every process concerning diffusion of a novel behavioural pattern (e.g. adoption of technology) as driven by imitation. Many economic models allow for alternative explanations for observed correlations in the timing of adoption by different economic agents (famous S-shape diffusion curves).[7] Still, other models consider the diffusion process as driven by imitative behaviour. We can divide them into models with direct payoff externalities, where imitative behaviour is driven by innate preferences for being similar to one's reference group,[8] and models with indirect payoff externalities, where agents have no innate preference for being conformists, but choose to imitate others for some other reason. The indirect externalities may arise for a variety of reasons: they may be a result of certain institutional arrangement as in the case of social norms (Elster 1989), competition for social status (Frank 1985), or, what deserves special attention in the light of social learning theory, informational externalities such as in informational cascades (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992).

2.3.2 Payoff externalities

Direct payoff externalities Gary Becker (1991) in his treatment of restaurant pricing describes an observation made at a popular seafood restaurant in Palo Alto, California:

> [Restaurant] does not take reservations, and every day it has long queues for tables during prime hours. Almost directly across the street is another seafood restaurant with comparable food, slightly higher prices, and similar service and other amenities. Yet this restaurant has many empty seats most of the time. [. . .] The same phenomenon is found in the pricing of successful sportive events, such as World Series and Super Bowls, and the related way in the pricing of best-selling books.

He notices that consumption of certain goods has not only private value, but also social meaning

> [A] consumer's demand for some goods depends on the demands by other consumers. The motivation for this approach is the recognition that restaurant

[7] For example, consider probit diffusion model (Davies 1979). In this model adoption of new technology happens without imitation. Actions of others have no effect on an agent. In fact, he even does not need to know what others are doing, because the only information he cares about is the threshold value.

[8] In the case of adoption by a firm, profit from adopting a technology might depend on the compatibilities between competing technical standards (Farrell and Saloner 1986).

> eating, watching a game or play, attending a concert, or talking about books are all social activities in which people consume a product or service together and partly in public. [...] [P]leasure from a good is greater when many people want to consume it, perhaps because a person does not want to be out of step with what is popular or because confidence in the quality of the food, writing, or performance is greater when a restaurant, book, or theater is more popular.

He proposes to make individual demand dependent of the aggregate demand (according to the neoclassical consumer theory it suggests that utility of consuming 'social good' depends on consumption choices of the others).

Becker is not the first economist to seek explanation for some anomalies in consumption of goods with social meaning by making the assumption of interdependence between individuals' demand schedules. Leibenstein (1950) employed a similar approach to formalize the verbal treatments of interdependencies in consumer choice due to Veblen and his theory of the leisure class, but also due to earlier writers such as John Rae and Pigou. He examined three types of effects that consumer choices have on individual demand: a *bandwagon effect* is an increase in individual demand due to popularity of the good, a *snob effect* is exactly the opposite, i.e. decrease in demand due to adoption of the good by other consumers, and the *Veblen effect* captures conspicuous consumption, an increase in demand for a good due to an increase in its price. He found that bandwagon and snob effects change the price elasticity of aggregate demand (with respect to individual demand), while Veblen effects may result in upward sloped parts of aggregate demand.

Extensive treatment of direct payoff externalities can be found in the literature on industrial organization, where they are referred as "network externalities" (Arthur 1989, Farrell and Saloner 1986). This kind of externality often appears in the context of a competition between different technical standards. Consider two standards (technologies) A and B, which are mutually incompatible. At each moment in time one potential adopter has to make his choice between the standards. Because of the incompatibility, a user of standard A is better off when he is surrounded by users of the same standard, A. Let the payoff to agent i from adopting a standard be the sum of the gains due to the technical efficiency of the standard and the gain due to compatibility with other users of the standard. We assume that the latter increases with the number of the users of the same standard. Suppose that *ceteris paribus* the standard B is more efficient then A.

Consider a situation where the difference in the numbers of users happens to be such that the gain due to the larger installed base offsets the loss from adoption of a technically inferior standard A. In such a situation a potential adopter should choose inferior standard A, thereby increasing the instalment base of standard A, and subsequently increasing the gain from adopting standard A even further. As a

result, the next adopter will also choose A, and so on: the system locks in an inferior (with respect to social welfare) state.

Examples of lock-ins to technically inferior standards are well known in the economics of technical change. The initial advantage of an inferior standard/technology that leads to suboptimal lock-in may arise for a variety of reasons. One such a reason often mentioned in the literature is related to first-mover advantage, and the most famous example of it is story of the QWERTY vs. Dvorjak keyboard (David 1985). Another factor that may cause an early lock-in to an inferior standard is related to sponsorship of a particular technology (Cowan 1990).

Payoff externalities may give rise to informal conventions. Consider right- vs. left- hand traffic. Once most drivers choose one of the alternatives, a driver would pay a dear price for choosing the alternative. As a result all drivers are better off taking the same side, even if for whatever reason some of them have intrinsic preferences for doing it the other way. Therefore we can say that one of the alternatives will be chosen by all drivers, and no driver would like to change her choice while others stick to theirs. The two choices correspond to two stable equilibria of a coordination game among perfectly rational individualistic agents. The problem of coordination can be solved by some informal agreement (Farrell and Saloner 1986).

Indirect Payoff Externalities Inserting externalities directly into the payoff (utility) function is a straightforward way to take into account social interactions and it allows for some insights into behaviour of a system with interdependencies between agents' choices. Nevertheless, one might feel uncomfortable about approaching the problem this way, because the question of why do agents imitate others has an extremely simple answer – they have preferences to do so! In some applications, e.g. competition of standards mentioned above, the origin of such preferences is transparent (whenever compatibility with other users is an advantage, it is better to adopt a popular standard) and the straightforwardness of the answer is well justified. In other applications, however, it is not clear how such preferences come about and further explanations are needed.

Payoff externalities may arise due to asymmetric information. For instance, Scharfstein and Stein (1990) explain mimetic isomorphism of new institutional theory mentioned earlier in a principal-agent framework. They examine a model with a population of investment funds managers of two types: "smart" ones who observe a signal correlated with the state of the market, and "dumb" managers who observe uncorrelated noise. The labour market can judge a manager only on the basis of information which consists of (a) profitability of investment, and (b) whether the investment behaviour of the manager is similar to behaviour of his peers. Notice, that in the absence of (b) smart managers' decisions would be correlated (since their signals are correlated with the state of market), while dumb managers would invest

at random. However, under (b) "herding" arises due to a "sharing-the-blame" effect: a manager who mimics the investment decisions of his peers is less likely to be blamed because then his decision is correlated with the decision of the others and this suggests to the labour market that he is likely to be smart. As a result even if a manager receives a negative signal about the market he may still decide to invest following the crowd to avoid being punished by the labour market. In Keynes's words "Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally".

As has been mentioned above, direct payoff externalities may result in the emergence of informal agreements (social norms)[9]. However many social norms are individually costly, and self-interested perfectly rational agents should not comply with such norms. For instance, consider racial (gender, cast etc.) discrimination. If, say, employers have positive taste for racial discrimination then they deliberately reduce the pool of potential employees and therefore an employer without racial prejudice should have competitive advantage (Becker 1976). In the long-run, in principle discrimination should disappear as a social norm even if no government/civil actions against racial discrimination are taken. Nevertheless, racial discrimination does exist, as do many other individually costly norms.

Another example of an individually costly social norm is a code of vengeance. Under certain circumstances (e.g. where corresponding formal institutions are weak or absent) the code may be socially beneficial, as it might work as a barrier to unlawful violence toward others. However punishing disobedience also involves costs: complying with a vengeance code a member of a family (clan, tribe etc.) has to take revenge, which means putting himself at risk. If this is the case, the party that is supposed to punish the deviator is better not doing it. Hence a rational decision is to leave it. Knowing this, rational agents would not follow the norm. But if everybody feels the same then revenge is not a credible threat, and vengeance cannot prevent violence. Nevertheless, a social norm that is individually costly may still be sustained if it is enforced by another norm.

Akerlof (1976) explained how such norms may be supported with the example of discrimination in the Indian caste system. While earlier analysis would conclude that discrimination disappears in the long-run due to room for arbitrage (Becker 1976), Akerlof showed that even individually costly customs may support themselves when not only breaking a norm is subject to penalty, but not punishing a rule-breaker is also punishable. He examined a simple model of a labour market under a caste system, in which employers have a choice between hiring labour according to caste codes with wage differentials, or according to output maximization regardless of

[9] In the context of imitative behaviour, social norms are interesting for two reasons. First, as is discussed above, social norms provide incentives to behave "as others do". Second, one of the main channels for diffusion of social norms is imitation.

caste. In addition, he assumes that a consumer who decides to buy goods from companies not using labour according to the caste code will be punished and become an outcast. He showed that the model has two equilibria: "low-trap" caste equilibrium, and no-caste optimal equilibrium. Once the system is trapped in the caste equilibrium, it may get out of it only if a significant share of agents decides to break the rules at the same time.

Notice that in this model the tendency to copy behaviour of others to comply with the socially imposed behavioural norm is not a part of innate preferences. Instead payoff externalities are introduced into the model in an indirect way via the institutional set up.

Although we examined only two economic contexts where payoff externalities arise even though not explicitly present in agents' preferences, there are many other instances where it happens as well (e.g. bank runs or financial crises). Now we proceed to a particular type of models with indirect payoff externalities inspired by the literature on social learning.

2.3.3 Informational externalities

Social learning theory outlined in section 2.2 emphasises "reciprocal determinism": environment via learning causes individual behaviour, but the individual behaviour, in turn, causes the environment. In the process of acquiring novel behaviour some knowledge is produced. It spills over to other agents and may induce some of them to copy the new mode of behaviour. This, in turn, also produces knowledge and enhances (or inhibits) the process of adoption. In contrast with the other models mentioned above, the process of adoption of a technology is driven solely by information externalities: agents' payoffs do not depend on the actions taken by others.

Bala and Goyal (1994) study a model of entry into a new market with unknown stochastic demand. They consider a pool of entrepreneurs who face a choice whether to enter the market or abstain from the entry. Decisions are made in sequential order, so that later entrepreneurs may observe the experience of their predecessors. They found that, first, if the pool consists of one entrepreneur, then with non-zero probability a market may disappear even if it is viable. They also explore the role of heterogeneity in entrepreneurs' beliefs: if beliefs are not heterogeneous enough then similarly to the case of single entrepreneur a viable market may be abandoned, while if the population is characterised by significant heterogeneity even non-viable markets never cease.

Caplin and Leahy (1994) examine a model of market crashes with firms trying to extract information about market viability from decisions made by other firms in this market. In their model entry into a market is two-staged: initial investment allows a firm to gather information about the state of final demand and on this basis it

decides whether to make additional investment and enter the market or abandon the project. Information gathered by a firm is private, however its decision (to entry into the market or cancel the project) is publicly observed. The information structure of the model implies that as far as the firms continue with their projects the knowledge of the market accumulated by firms is effectively "trapped" in private hands: there is no way to distinguish between the behaviour of a firm that has received good news and a firm that has received negative information. The situation changes dramatically when some firms decide to suspend their projects. Suspensions release negative information and the market suddenly collapses.

Bolton and Harris (1999) examined a model of social learning based on a game of strategic experimentation. Their model is an extension of the well-known two-arm bandit problem. One of the bandit arms (the "safe" arm) provides some known payoff, the other generates random payoff with unknown distribution (the "risky" arm). The bandit problem is a classic example of the trade-off between experimentation and exploitation. While usually bandit models examine a player who makes his decision in isolation based on information from his own experience with the machine, Bolton and Harris consider a population of players each of whom face the same choice between "safe" technology with known payoff and "risky" technology with unknown payoff. Time is continuous and switching between technologies is costless. Players dynamically allocate their time between the technologies. Payoffs are publicly observable.

They prove existence of a symmetric equilibrium and prove its uniqueness.[10] They show the presence of two motives in agent's behaviour working in opposite directions. First, there is the free-rider effect: a player may be tempted to devote time to the known technology, in the hope of free-riding on the experimentation performed by others. Second, there is the "encouragement effect": a player may choose to experiment, in order to encourage experimentation with the risky technology by others, in order to benefit from the induced positive informational externality.

Informational cascades Informational cascades represent a particular case of models for social learning that became popular among economists after the seminal works of Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Welch (1992). Let us follow the presentation of Bikhchandani, Hirshleifer, and Welch (1992) model given in Bikhchandani, Hirshleifer, and Welch (1998).

Consider a population of agents presented with a choice between two actions. In the context of technology adoption it may be *Adopt* the technology or *Decline*. The payoff to adoption is either +1 or -1 depending on the state of the technology: "High" or "Low" respectively. Without any further information both states of the technology have equal probabilities (i.e. prior probabilities are 1/2 for both "High"

[10] Although other non-symmetric equilibria might exist.

and "Low" states of the technology).

Agents make their decisions in sequence; the order in which agents decide is exogenous and known to all agents. Information about a technology comes to an agent from two sources: a private signal and information about actions of her predecessors. Private signals are conditionally independent and may be of two types "High" and "Low". Probabilities to receive a particular type of signal depends on the state of the technology, when in the favourable state of the technology the probability to receive "High" (p) is higher than the probability to receive "Low" (1-p), while when the state of the technology is "Low" the probability of a "Low" signal is higher then the probability of a "High" signal. For simplicity we assume that the precision of a correct signal is the same for "High" and "Low" states of the technology. Table 2.1 presents conditional probabilities for the signals. Notice, that if agents' decisions were based solely on private information, then agent's optimal strategy is to "follow the signal" i.e. to adopt if the signal is favourable, otherwise reject the technology.

Table 2.1: Signal probabilities

	$\Pr(v = H\|V)$	$\Pr(v = L\|V)$
$V = H$	p	$1 - p$
$V = L$	$1 - p$	p

The other source of information is the history of adoption. Had all agents followed their signals, agents' private information would be revealed to the public. However it is not always individually optimal to follow one's own signal. At a certain point positive (negative) public information exceeds a threshold and starts to outweigh private signals. Then, instead of taking action in accordance with private information, the agent should "follow the herd", and adopt (reject) the technology regardless to his private information. *An informational cascade is said to occur if an agent's action does not depend on her private signal* (Bikhchandani, Hirshleifer, and Welch 1992).

Once an informational cascade starts actions of all further agents become uninformative. Indeed, after the numbers of adopters and non-adopters are such that agent i has to follow the herd, the next agent, $(i + 1)$ must infer that i discarded private information. Hence i's action reveals no new information to $(i + 1)$. As a result, $(i + 1)$ finds herself in the same situation as i, and therefore must ignore her private information as did i, and so does $(i + 2)$ and so on. Thus when an informational cascade occurs, accumulation of public information ceases and conformity arises.

Several features of the model are worth mentioning here. First, notice that similar to other models with lock-in, an action widely adopted does not have to

be the "correct" one. Indeed, even if the state of the technology is "High", an unlucky chain of events (negative signals) might result in widespread rejection of the technology (and vice versa an informational cascade might lead to adoption of a mediocre technology). Second, notice that were agents able to observe *signals* rather than *actions* of their predecessors the true state of technology would be figured out with probability 1. Existence of suboptimal lock-ins in the model is due to the restrictions on publicly available information. Third, cascades are "fragile" [11]: the arrival of better informed individuals, the release of new public information, or changes of the underlying value of adoption could dislodge an informational cascade. A cascade emerges when agents acting in self-interest find it optimal to follow the herd and do not reveal their private information. If, however, in the course of the process an agent decides to follow the signal and in contravention to popular behaviour, then her action would reveal her private information and unlock the cascade. Bernardo and Welch (2001) have shown that irrationally overconfident entrepreneurs who overweight their private signals with respect to signals of the others enhance social welfare.[12]

2.3.4 Bounded rationality

Although the assumption of perfect rationality of human beings has allowed economists to build parsimonious yet self-coherent models of individual behaviour, even most vehement proponents of the neoclassical approach do not argue that humans are rational to the degree their models prescribe. To address the problem a variety of models that account for bounded rationality of individuals have been proposed. It is impossible to give any complete review of them in a short space; instead here we focus on several key references on social learning with boundedly rational individuals.

Kirman's (1993) work on "ants, rationality, and recruitment" seeks to explain the asymmetric behaviour of market participants in an apparently symmetric situation by analogy with the process of the recruiting behaviour of ants. In an experiment by entomologists a colony of ants was allowed to feed on two identical sources of food. Interestingly they documented that the pattern of exploitation of these sources was highly asymmetric: while intuitively one might expect that in the long run ants would be split in half between the sources, if fact they stabilized with 80 percent at one source, and the rest 20 percent at the other. Moreover, from time to time "flips" occur between the concentrations at the two sources, i.e. a source that was exploited by some 80 percent of ants before the "flip", attracts 20 percent after the "flip". An analogy with traders' behaviour is obvious.

[11] This feature differs them from other models with positive feedback

[12] They examine an evolutionary group selection model and found that groups with overconfident entrepreneurs have higher chances to survive.

Kirman formulated a simple model of such asymmetric behaviour based on a particular form of a Markov chain similar to Polya urn processes. Consider two kinds of stocks[13] available at the market, say stocks A and B. Some traders prefer to hold stock A, while others prefer B. The state of the system can be characterized by the "market shares" of the stocks.

Each period a trader meets another trader chosen at random from the population. With some probability the first trader adjusts his preference to the preference of the second (due to the symmetry it makes no difference who is the first and who is the second), i.e. changes his preference, if the other one has different preference, or keeps her view if the second trader is of the same type. There is also a small probability ϵ that the first trader changes his view independent of meeting the "model". In the real world it might happen, for instance, due to arrival of exogenous "news", or replacement of the existing trader by a new one with the different view.

Depending on the parameters of the model the *equilibrium distribution* of A and B shares, the fraction of the time the system spends in each of its states (particular splits of the market), is U-shaped, flat, or inverted U-shaped.

An asymmetric equilibrium distribution (U-shape form) that corresponds to the observed pattern of ants' (and stock market) behaviour arises when the parameter ϵ is small enough, i.e. the effect of social interaction overwhelms exogenous influence. In this case the system spends most of the time near one of the boundaries, i.e. when the one of the stocks is much more 'popular' than the other, with occasional 'flips' in traders' preferences.

Several features of the model are particularly interesting in our context. First, there is a sharp contrast between the complex dynamics of the system (resembling stock market 'sunspot' investment targets), and yet the very simple behaviour of agents. It highlights the importance of social interactions that have to be taken into consideration if one is to build any realistic model of human (and apparently ant) behaviour: "the behaviour of the group as a whole cannot be inferred from analysing one of the identical individuals in isolation. Without taking explicit account of interaction between individuals, the group behaviour ... cannot be explained." (Kirman 1993, p.137)

Second, although the process of decision making is not explicitly specified in the model (hardly any economist may claim to know the ants' map of preference), the behavioural rule is consistent with the social learning theory (discussed in the section 2.2.1). Indeed, the second trader can be considered as a "model" for the first one. Through observational modelling the first trader may adopt the behaviour of the "model" and, in turn, pass it to another trader and so on and as a result a new pattern may arise.

The paper does not explicitly discuss the implications of imitative behaviour for

[13]The exact number of alternative stocks is not important

overall social welfare. For that, let us assume that underlying an agent's behaviour rule there are some (perceived) payoffs and those payoffs are different for different actions, say, stock A has higher returns than stock B, and they are reflected in the probabilities of 'conversion', i.e. preference for stock A is more likely to be 'imitated' than preference for stock B. This difference in transition probabilities would bias the equilibrium distribution toward stock A. However, the general feature of the distribution, its U-shape, would still hold. Therefore, even though the system would spend more time in the socially optimal state (with most agents holding stock A), there still would be a non-zero probability of a temporary "madness of the crowd", when all of a sudden the system will get stuck in the sub-optimal state.

Also notice that in the model the 'society' is not structured: agents meet each other at random[14]. At the same time, in many situations the structure of contacts in the population is highly relevant for understanding the dynamics of the process. An obvious example is the spatial dimension of the process. If interactions between agents have a local nature, we are likely to see localized emergence of behavioural patterns, which may differ from place to place (e.g. consider left- and right- hand traffic).

Ellison and Fudenberg (1993) examined "rules of thumb" in social learning. As well as Kirman (1993) in their model(s) agents follow some fixed behavioural rules ("rules of thumb"). The rules of thumb they study try to capture the process of social learning through "popularity weighting": agents (individual or firm) incline to use technology that did best in the previous period. Drawing an analogy with imitation studies in organization theory one might say that the model blends frequency-based imitation mode as reflected by popularity weighting, and outcome-based imitation driven by current payoff difference.

The structure of the simple model with a homogeneous population is as follows. There is a continuum of agents who can (and must) choose between two technologies. Each period only a fraction of the population can rethink its choice. The difference in payoffs to adopting technologies at time t is a random variable that is a sum of some unknown constant and a random shock. An agent can observe adoption decisions and payoffs associated with the decisions. The agents' decision rule, the rule of thumb, is based on "popularity weighting": an agent chosen to revise her technology choice at time t takes into account both current payoffs (realization of payoffs in the previous round), and popularities of the technologies. The rule of thumb is characterized by a parameter that describes "popularity weight" vs. current payoffs difference in the decision making.

Ellison and Fudenberg (1993) examined a particular case where random shocks are distributed uniformly. They asked whether the system converges to one of its boundary states (i.e. when one of the technology dies out), and under which

[14]It corresponds to fully-mixed approximation in epidemiologic models of contagion.

parameters.

They find that depending on the weighting parameter, three possible situations may arise. First, there may be an "optimum" weighting when for any initial condition the system converges with probability one to the better technology.

Second, there may be "overweighting" when the system always converges to a steady state. However in contrast with the case of "optimum" weighting, whether the better technology will be selected depends on the payoff difference and the initial shares of the technologies. When expected payoff difference is high, then with probability one the better technology will be selected, regardless of the initial condition (i.e. the behaviour of the system is very similar to "optimum" weighting case). When the difference is not high enough to ensure convergence to the better technology, then all depends on the initial conditions: if the initial share of a technology is high the system converges to the technology that was more popular at the beginning of the process (path-dependency), otherwise one cannot say in advance which of the two technologies will prevail in the long run (both steady states have positive probability).

Finally, the third case when agent's decision is based predominantly on the current payoffs (the value of the weighting parameter is low) is the situation of "underweighting". In this case the system may not converge at all. Similarly to the case of "overweighting" the difference in expected payoffs is crucial for prediction of long run behaviour of the system. If the difference in expected payoffs is high enough, then with probability one the better technology will be selected regardless of the initial condition. Otherwise the system has a nondegenerate invariant distribution.

Taking a broader perspective we can draw parallels with the organization theory literature on imitation and make conjectures about the diffusion path of competing technologies in relation to organizations' strategies/"behavioural traits". In the environment where agents put enough weight on the most popular choice (frequency-based imitation) the process of social learning always leads to conformity. On the contrary, when the decision about technological choice is heavily affected by (contemporary) performance of the technologies (outcome-based imitation) the system never settles down and both technologies will be in use at any time (on management fashion cycles see a model by Strang and Still (2004)). Moreover, a large enough payoff difference ensures that the system converges to better technology no matter where it starts. In the situation when popularity weighting is overwhelming ("overweighting") but the difference in the payoffs is moderate, the system may converge to a suboptimal steady state.

Speed of diffusion of a new superior technology in their model depends on the payoff difference. This result is consistent with empirical findings from diffusion research. Indeed, according to empirical evidence on the diffusion of innovations profitability of an innovation increases the speed of its adoption (Griliches 1957, Mansfield 1961). Another interesting finding is that for a fixed payoff difference, the

speed of convergence decreases as the magnitude of the random shock is increasing. Haunschild and Miner (1997) examined the decision of hiring of an investment banker to advise an acquiring firm on an acquisition, and found that uncertainty is likely to increase the probability of frequency- and trait- based imitation and to decrease the probability of outcome-based imitation.

Ellison and Fudenberg (1995) modelled social learning in an environment where information spreads via "word-of-mouth". Similarly to Ellison and Fudenberg (1993) there are two technologies with unknown payoffs, and each period only a part of the population may reconsider choices. Payoffs to adoption of the technologies are subject to two kind of shocks: common shocks shared by all agents and idiosyncratic ones specific for each of the agents. "Word-of-mouth" is introduced into the model through random sampling of the population: an individual who is to reconsider her choice makes a random sample of N (exogenous) people and adopts the technology with the highest payoff. The main focus of the study is to find the conditions under which the system exhibits conformity and under which it is characterised by diversity.

First, they find that conformity is achieved if there is *not too much communication*. Furthermore, socially efficient outcomes, i.e. convergence toward the technology that does better on average, can be achieved only when there is very little communication. That might seem to be somewhat surprising, especially if compared with the results of their previous model discussed above. Indeed, while the "rule of thumb" model says that lock-in to one technology (conformity) is impossible without sufficient social learning expressed in popularity weighting, the "word-of-mouth" model suggests that social learning tends to push the system away from a lock-in.

The reason that social learning has to be restricted in order to let the system settle down is the 'must-see' condition in the later model: an agent can adopt a technology only if she sees it in her sample. With this condition unpopular technologies might die out: if the size of agents' samples is relatively small the unpopular technology will not be present in the samples of many thus its share will decrease over time until it will eventually disappear. On the contrary, if the size of the sample is not small even the unpopular technology will end up in the samples of many and under some realization of random shocks it might have a dramatic comeback.

The second result is related to the efficiency of the social learning: for convergence to the better technology the intensity of "word-of-mouth" (sample size) has to be small, but not too small. Indeed, as we know when the size of the sample is not small enough, the system never settles. If however the sample is too small the system has excessive inertia such that an unlucky realization of random shocks might lock the system into the suboptimal choice. In the middle there is a region of sample sizes where the system on the one hand may settle down, but on the other, due to the fact that on average one technology is better than the other, it avoids the suboptimal lock-in.

2.4 Four themes

In the rest of this chapter I will focus on several economic contexts for social learning and imitation to be explored in detail throughout this dissertation. We start with the relationship between the structure of the communication network and the efficiency of social learning. As information and communication technologies (ICTs) are changing the structure and intensity of information flows in the society, there is a concern that although, on the one hand such ICT-based globalization enhances welfare via exchange of information about more effective technologies, practices and ideas, on the other hand it may increase the probability of a 'global lock-in'.

Another context discussed in this dissertation is related to financing a new technology. There I will argue that there is a need for a model of social learning that explicitly takes into account the significant uncertainty surrounding the development of the market for a new technology, the limited scope of publicly available information, and the asymmetry in the distribution of private information about the technological value and market perspectives of the technology.

While two themes of the dissertation mentioned above are directly related to social learning process, the other two are concerned with some implications of social learning and imitation. Following Veblen's line we analyze effects of conspicuous consumption on the diffusion of a new good. Consumption patterns differ among classes, and inter- and intra- class imitation of these patterns is constrained by social norms. We enquire about the role played by class structure and social norms in the process of diffusion of a novel product with characteristics of a status good.

The last context studied in this dissertation concerns the relationship between path-dependency in the process of technical change and the value of innovations. The underlying practical question is how one can explain the observed distribution of innovation values. We will argue that to explain this distribution one has to take into account some important characteristics of the process, such as clustering of R&D efforts in technological space (one reason for which may be imitation and social learning).

2.4.1 Social learning in networks

Recent major advances in information and communication technologies, sometimes referred as the IT (or ICT) revolution, seem to have radically changed our lives, and promise even more change in the years to come. One can 'travel' to places, one never knew before, not leaving the couch in front of one's TV, or learn an amazing variety of languages or cooking styles just from the computer on one's desk. Although the benefits of these innovations appear to be obvious, we may wonder whether this increasing intensity of information flows is always a good thing.

Let us try to project history into the future. Globalization is not something

absolutely new and unique for the age of IT. Indeed, in the magnitude of the effects of globalization perhaps we are yet to reach the time of the Great Geographical Discoveries, which shaped the world as we know it today. On the one hand, one could say that mankind as a whole benefited from those discoveries in the sense of bringing new, more productive technologies to the places where they had not been used before, to name only a few - the wheel and the horse to the Americas; corn and the potato to Europe. But, on the other hand, it came at a price, and this price was paid mainly by local populations. The new technologies required land and labour, hence the land and labour were expropriated from locals. Moreover, wherever the local supply of workers was inadequate for large-scale production slaves were imported. Thus, while new technologies were more efficient, the local populations were excluded from the benefits of the increased productivity.

One may argue that things are different now. In contrast with the early days, 'ICT globalization' is based not so much on the exchange of material goods, but rather on the exchange of information. In a 'knowledge economy' technologies are no longer embedded in capital goods, but rather in virtual goods (information). If so, then more productive technologies can be transferred worldwide, and it will not be easy and perhaps impossible to exclude somebody from the profit generated by his knowledge. Does it mean that the 'information-based' globalization is going to make us better off, while at the same time letting us avoid the worst consequences?

In a recent series of rigorous and extensive studies Bala and Goyal propose that the answer is likely to be '*yes*' (Bala and Goyal 1998, Bala and Goyal 2001),[15] because 'better practices' always win in the long-run (under some conditions, to be discussed in the next chapter). Acknowledging the importance of their work, we shall nevertheless argue that the prospects of IT globalization might be not so charming if one takes into account the possibilities of a "lock-in", as emphasized by the students of economics and history of technical change (David 1985, Cowan 1990).

In the next chapter we examine a modification of the model of informational cascades (Bikhchandani, Hirshleifer, and Welch 1992) where similarly to Bala and Goyal (1998) we assume communication structure in the society – agents may observe only their direct neighbours. Our main interest is the relationship between the scale of communications, expressed as network density, as well as the structure of the communication network on the one hand, and social welfare on the other.

2.4.2 Financing new technologies

The last decade of the previous century was filled with superoptimistic expectations about the new economy. Investment in technology ventures was considered a must, finding a business concept based on the Internet seemed the only way to survive

[15] This is my interpretation of their results

in the upcoming age of the new economy. At the height of the dot-com frenzy a computer journal stated (Computerworld, June 1999)

> Companies raid one another for employees, paying even better salaries and dangling ever-more stock options, and workers tell of the four companies they have joined in as many years. More money than ever is flowing into technology venture capital funds. Established executives in old-line businesses are quitting to join the rawest start-ups, and young people flock to the [Silicon] Valley in the faint hope of creating or being part of a start-up company. Everyone hopes to become another eBay.

Just a year later, the mood changed to exactly opposite. The same investors chasing internet start-ups only several months before suddenly decided to withdraw funds from most dotcom business ventures. The minds of industry observers were preoccupied with somewhat different questions (FT, August 2000)

> How did they ever get financed? Well, ... nothing destroys value like underpriced capital. The dotcom bubble will be studied not so much for its technical or entrepreneurial innovation, but as a capital market anomaly. [...] When capital is nearly free, the need to discriminate between new ideas disappears.

What is interesting in this story from the point of view of social learning is that both sides of the market: the entrepreneurs founding dotcom start-ups, as well as investors pouring money into these ventures, happened to be overoptimistic about the future of Internet economy.

Taking a more general perspective on financing a new technology with uncertain market prospects one may notice that the information on which market participants base their expectations is segmented, and there is asymmetry in the distribution of this information between the two sides of the market. The innovation management literature distinguishes two main sources of risk in the process of development and commercialization of a new technology. *Technical risk* is the risk of "failure in the attempt to convert invention to innovation (as when the product does not meet specifications in terms of performance, cost of production, or reliability)." (Branscomb and Auerswald 2001, p.4.) *Market risk* arises due to uncertainties about the demand for the product that might not be formed yet, and may include the risk " that the product will [not] provide vectors of differentiation sufficient to distinguish it from competitive offerings" and the risk "that the proposed business model will [not] be successful in the market." (Branscomb and Auerswald 2001, p.4.)

A distinctive feature of these markets is the asymmetry in the distribution of information about those risks. With reference to the two sides of the market:

> On each side of the Valley of Death stands a quite different archetypal character: the technologist on the one side, and the investor/manager on the other.

> Each has different training, expectations, information sources, and modes of expression. The technologist knows what is scientifically interesting, what is technically feasible, and what is fundamentally novel in the approach proposed. [...] The investor/manager knows about the process of bringing new products to market, but will likely to have to trust the technologist when it comes to technical particulars of the project in question.[...] To the extent that technologist and investor/manager do not fully trust one another or cannot communicate effectively, the Valley of Death between invention and innovation becomes deeper still." (Branscomb and Auerswald 2001, p.12)

The "information and trust gap" has received significant attention in the recent years because of the important role which small enterprises play in rapidly growing high-tech industries. In particular, much research effort has been focused on understanding how the problem of asymmetric information affects the mode of financing (multistage investment, VCs' syndication, etc.), exit strategies, venture capital structure etc. (Gompers and Lerner 1999, for a review). However, so far the unit of analysis in this field has been the single VC-firm relationship, and the literature has yet to study intra-population effects (e.g. interdependencies between the VCs' choices in terms of technologies).

Models of social learning seem to be adequate tools for studying of how 'information and trust gaps' affect behaviour of entrepreneurs and investors in such markets. This approach should not be considered as a substitute, but rather as complementary to other models for financing new technologies based on principal-agent theory mentioned above, because of the difference in the level of analysis.

However to take into account the peculiarities of the market the models have to be extended in several ways. First, two-sided interactions are to be modelled explicitly, for example, as a non-cooperative game. Second, we also shall assume that the success of a venture depends on both the quality of the technology and its market prospects and that information about the technology and the market is limited by the corresponding sides of the market: entrepreneurs know the technology, while investors have better knowledge of the market. Finally, we shall assume that public knowledge consists of the outcomes of negotiations rather than actors' actions.

2.4.3 Conspicuous consumption

Early economists and contemporary economic historians alike have identified conspicuous consumption as an important determinant in the expansion of markets and technological innovations in the Western world in the 18th century and later. The acquisition and diffusion of consumer goods is driven by "the recognition and admiration of our fellow human beings", as "to deserve, to acquire, and to enjoy, the respect and admiration of mankind, are the great objects of ambition and emulation", (A.Smith, cited in (Rosenberg 1968, p.365)).

There is interdependence in the choices of the different populations of adopters. Members of different social groups observe the consumption patterns of other members in society. In the absence of more direct social contact consumption patterns reveal the social status of people. This is a process in which consumers compare, evaluate and imitate or reject the choices of relevant others.

Furthermore, socio-economic attributes such as disposable income or more pervasive value systems have an influence on the choice and the subsequent legitimization of an innovation in consumer goods. The diffusion of consumer good innovation does not just involve the dissemination of information as the diffusion literature typically would suggest (for an overview see Geroski (2000)), but is determined by a social process of persuasion and depends on the extent of consumer heterogeneity in an economy.

Over the years the interest in the study of consumer behaviour under the presence of externalities amongst consumers has steadily increased. Many contributions have drawn on ideas set out in the classic works of John Rae (1905) and Thorstein Veblen (1921) on conspicuous consumption as well as on ideas by sociologists such as Georg Simmel (1957) or Pierre Bourdieu (1984). Most of this literature has addressed the allocation aspect of interdependent preferences and studied the adoption of pure luxury goods, such as fine art, holiday resorts, luxury cars or fashion goods (Pesendorfer 1995, Swann 2001, e.g.).

Perhaps the focus on the consumption of luxury goods may explain why diffusion of new positional goods is rarely studied in this literature. Meanwhile, understanding what factors may explain the pace with which a market for new products grows over time is essential for understanding consumer goods innovation. It is often the case that new technologies are developed by companies in anticipation of fast growing demand from consumers. A fast growing market attracts more companies, and to sustain the competition they have to invest more and more in R&D. One can argue that this was the case with the market for automobiles in the beginning of the 20th century, plastics in the middle of the 20th century, and it seems to be an essential factor in the development of wireless communication technologies at the end of the 20th century.

Conspicuous consumption certainly plays an important role in the process of the diffusion of new consumer products, and it is well known to manufactures. They exploit it to extract more profit from a single innovation. Introduction of new products to the market is often made in steps: first, companies launch a new product at the top price, at the top quality end of the market, so that possession of the product can be considered as a credible signal about social status. Later, they dilute the status of the good, introducing new models of the product at a lower price, increasing demand and exploiting the economies of scale in mass production of the good. Therefore, it seems to be interesting to know the factors that affect the speed of growth of demand for new products in the presence of conspicuous consumption

effects.

Some research has partly addressed this question. For example, it has been shown that social norms that determine the strength of the desire for distinction or conformity are essential elements in the dynamics of demand. Cowan, Cowan, and Swann (1997) have devised a stochastic model whose dynamic is based on aspiration-, bandwagon- and Veblen effects. They show that if certain consumer groups seek distinction and others aspire to their behaviour, cyclical consumption patterns and consumption waves may emerge. In a similar fashion Janssen and Jager (2001) explain market dynamics with lock-in, fashions or unstable renewal. They posit that the dynamics is dominated by the behavioural rules of consumers reflecting their preference for either distinction or conformity.

Although those models are based on the fact that agents reference groups are determined by agents' belonging to certain class (or social status), they do not study the effects of class structure on diffusion paths. Nevertheless, one might expect that class structure plays a significant role in determining both speed and penetration level of diffusion. To examine this idea, in Chapter 5 we build a simple evolutionary model to analyse the effects of social norms and class structure on the characteristics of diffusion.

2.4.4 Path-dependency of technical change and the value of innovations

It is well recognized these days that only efficient production, accumulation, and utilization of technological knowledge can ensure long term economic growth. Planning and implementing R&D programmes have become a routine task for many governments and companies around the world. Therefore knowledge about the distribution of returns from R&D is of great practical importance.

The main problem hindering research in this direction has been scarcity of data on R&D. However, with the arrival of new data, particularly patent data, and with advances in methodology the field is rapidly expanding. The evidence accumulated in recent years confirms earlier findings and are univocal on the overall features of the distribution of the innovation values: it is highly skewed with most of the innovations having value close to zero, and few innovations scoring very high, a fact that has direct implications for planning and evaluation of innovation policies and firm strategies (Scherer and Harhoff 2000).

Although the extreme skewness of the distribution is now a well established fact, the precise form of the distribution of the innovation values is still under debate. In particular, there is a controversy about the right tail of the distribution. Based on the results of a survey of holders of German patents Harhoff, Scherer, and Vopel (1997) report that the best fit for the tail (defined as innovations with values over DM

23,000) is obtained with a lognormal distribution (vs. Pareto and Singh-Maddala distributions). On the contrary, applying techniques of extreme-value theory to the set of different data on the innovation values, Silverberg and Verspagen (2004) demonstrate that if the lower bound of the tail is set correctly, the tail is fitted better with a Pareto distribution (than with a lognormal distribution).

The question about the tail of the distribution of the innovation values is relevant not only for applied policy analysis (Scherer and Harhoff 2000), but also for more theoretical research. Kortum (1997) examines a search-based growth model and shows that exponential growth can be achieved only if the distribution of innovation sizes is Pareto-like. In a recent paper Jones (2005) extends the analysis by Houthakker (1955) with respect to the microfoundation of a production function. In his model, the global production function has the shape of the familiar Cobb-Douglas specification when ideas are distributed according to Pareto. His model also has strong predictions for the direction of technical change: technical change is purely labour-augmenting in the long run.[16]

So far, research in this direction has been focussed on the properties of the distribution. I propose to approach the problem from the other end: instead of questioning what is the exact form of the distribution of innovation values we inquire about the *process* that generates the distribution. As we will see the evolutionary theory of technical change is helpful in understanding the dynamics of innovation values.

The argument proceeds along the following lines. According to evolutionary theory the process of technical change is incremental and path-dependent, and the development of a technology follows "technological trajectories". Success of an innovation, related to the resolution of an important design problem, plays the role of "focusing device" (Rosenberg 1969): it directs innovative search to the areas of "technology space" opened by the innovation, and stimulates the flow of inventions based on the technology it represents. Following this logic we make an assumption that the value of the innovation depends on the range of the problems it can be applied to: the more general it is, the higher is its value. Combining this assumption with path-dependency in the process of technical change we expect the dynamics of innovation values to be path-dependent: the more valuable the innovation is, the more likely it is to be employed in consequent innovations. As a result, the more valuable it will become.

There are (at least) two factors behind path-dependency in technical change which tend to 'bunch' technologies together: (a) complementarities between contemporary technologies, and (b) localization of the search in the technological space, due to the bounded rationality of agents.[17] More detailed discussion of these fac-

[16]In what follows I will use different notion of the direction of technical change: it is the direction in "technological space" rather than long-run changes in capital and labour shares.

[17]For different technologies the relative importance of systemic and cognitive factors mentioned

tors will be given in Chapter 6. Here we only highlight the relationship between path-dependency and imitation.

First, it is worth emphasising that given our research question we shall view the search process not at the level of individual agents performing their search on their own, but as a process that involves the whole technological community; this community includes inventors, firms, government labs, academicians and the like. There is an obvious parallel with the sociology of science, in particular, with Thomas Kuhn's "scientific paradigms". During the stable phase of the development of a technology researchers and engineers have a number of standard approaches to solve standard problems shared by the community. To solve a particular engineering problem means finding an appropriate standard solution (design) and adjusting it to the problem (Cooper 2000).

Furthermore, the *research agenda* (i.e. what needs to be improved, what can be achieved with available techniques *etc.*) is also shared at the community level. As a result, the direction of innovative search is framed by the current state of the technology and hence depends on the previous success (in terms of both technological achievements and commercial benefits). Such a picture of innovative search goes along with the views of Rosenberg (1969, 1974) who sees inventive activities as focused on a set of related engineering problems ("focusing devices/technological imperatives") which result in "compulsive sequences" of innovations over time.

2.5 Summary

Imitation is not peculiar to human beings, but rather common among other animals. However it is our species whose survival in the course of evolution has been primarily dependent on the (high) propensity for imitation and social learning. Therefore few models on animal learning may be helpful to understand human behaviour. It should be stressed that imitation is not limited to the level of individuals, but in addition the behaviour of groups of individuals, an important example of which are economic organizations, can be described in terms of imitation and social learning as well. The main focus of the following essays will be on imitation and social learning in a few economic phenomena. We will examine effects of social learning in some particular settings characteristic to economic environment exploring in more detail the four themes outlined in this chapter.

here may differ.

Chapter 3

IT Revolution, Globalilzation and Informational Lock-In

3.1 Introduction

Recent major advances in information and communication technologies, sometimes referred as IT (or ICT) revolution, seem to have radically changed our lives, and promise even more change in the years to come. One can 'travel' to the places, one never knew before, not leaving a couch in front of one's TV, or learn an amazing variety of languages, or cooking styles just from the computer on one's desk. Although the benefits of these innovations appear to be obvious, we may wonder whether this increasing intensity of information flows is always a good thing.

Let us try to project the history into the future. Globalization is not something absolutely new and unique for the age of IT. Indeed, in the magnitude of the effects of globalization perhaps we are yet to reach the time of the Great Geographical Discoveries, which has shaped the world as we know it today. On the one hand, one could say that the mankind as a whole benefited from those discoveries in the sense of bringing new, more productive technologies to the places where they had not been used before, to name only a few - the wheel and the horse to Americas; corn and the potato to Europe. But, on the other hand, it came at a price, and this price was paid mainly by local populations. The new technologies required land and labour, hence the land and labour had been expropriated from locals. Moreover, wherever the local supply of workers was inadequate for large-scale production slaves were imported. Thus, while new technologies were more efficient, the local populations were excluded from the benefits of the increased productivity.

However, one may argue that the things are different now. In contrast with the early days, 'ICT globalization' is based not so much on the exchange of the material goods, but rather on the exchange of information. In a 'knowledge economy'

technologies are no longer embedded in capital goods, but rather in the information. If so, then more productive technologies can be transfered worldwide, and it will not be easy if not impossible to exclude somebody from the profit generated by his knowledge. Would it mean that the 'information-based' globalization is going to make us better off, while at the same time letting us avoid the worst consequences?

In a recent series of rigorous and extensive studies Bala and Goyal (referred as BG henceforth) propose that the answer is likely to be '*yes*' (Bala and Goyal 1998, Bala and Goyal 2001),[1] because 'better practices' always win in the long-run (under some conditions, as to be discussed later). Acknowledging the importance of their work, we shall nevertheless argue that the prospects of IT globalization might be not so charming, if one takes into account the possibilities of a 'lock-in' emphasized by the students of economics and history of technical change (David 1985, Cowan 1990).

Although most models of 'lock-in's assume some kind of 'network effects', i.e. payoff externalities (Arthur 1989), there are some which consider only effects of information (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992)[2] . To make the argument we have chosen the model of Bikhchandani, Hirshleifer, and Welch (1992) (BHW henceforth). Our main interest is the relationship between the scale of communication, expressed as network density, as well as the structure of the communication network on the one hand, and social welfare on the other.

The rest of the paper is organized as follows. In the next section we will review the basic BHW model of informational cascades, and agents' decision rule. Then we will formulate and examine a model of informational cascades in a network. Discussion of our results follows in Section 3.4. The last section concludes.

3.2 BHW model

There are several models of social learning and imitation besides BHW, discussed in Chapter 2, but BHW has some features which makes it attractive for our purposes:

(i) The model is based on *individual rationality*. Whether this concept is realistic or not remains beyond the scope of this chapter (for discussion see section 2.3.1), but at least it exempts us from making too many *ad hoc* assumptions about agents' motivations, decision processes, role of social interactions and so on.

(ii) The model captures (although very schematically) both emergence and diffusion of behavioural patterns/"common practices". It is essential that the model allows for a lock-in to (*ex-post*) suboptimal outcomes.

[1]This is my interpretation of their results

[2]Presence of 'network effects' is likely to strengthen our results

(*iii*) The model does not require superrationality from our agents. The decision rule is simple, transparent, and intuitive.

(*iv*) The model does not require superrationality from us either. It is easy to handle and allows us to generate treatable results with fairly modest means.

On the top of that, the predictions of the model have been proven to fit with the results of experimental studies (Anderson and Holt 1997). Below we describe the BHW model as it was presented in Bikhchandani, Hirshleifer, and Welch (1998).

There are n identical agents each of whom has to choose between the two actions: to switch to a new 'practice' (adopt the new technology), or to keep *status quo* (reject the technology). The payoff to adopting the technology, V, is either 1 or -1, with equal probability. The payoff to keeping the *status quo* is 0. BHW assume *sequential* decision making, such that one (and each) period only one agent makes his choice. The order in which agents decide is exogenously given and known to all.

The information available to an agent includes both a (private) signal and public information which consists of the history of his predecessors' decisions. The agent's signal, v_t ($t = 1..N$), is either $High$ or Low. If the technology is worth adopting, i.e. $V = +1$, the probability to receive a $High$ signal is $p > 0.5$, while if $V = -1$, the probability of $High$ is $(1 - p)$. The conditional probabilities of receiving signals are summarized in Table 2.1. The precision of the signal, p, is the same for all agents. The signals are identically distributed and independent conditional on V. Agents are risk neutral and choose the action which has a higher expected payoff. If an agent is indifferent between the two alternatives we impose a tie-breaking rule somewhat different from the original BHW setting,[3] namely we assume that in case of a 'draw' he puts slightly higher weight on his private information and follows his signal. This kind of tie-breaking rule was used by Anderson and Holt in their experiments (Anderson and Holt 1997).

Details for the solution of the agent's problem can be found in Appendix A. The agent's decision rule is as simple as this: if by the time the agent is to take the decision the difference between the number of the agents who adopted (adopters) and the number of the agents who rejected the technology (non-adopters), d, is greater or equal to +2, then he adopts regardless of the private signal; if the difference is less or equal to -2, he should reject whatever his signal is; if neither is the case, then he follows the signal, i.e. adopts when the signal is $High$ and rejects when it is Low.

BHW define an *informational cascade* as a situation where the decision of the agent does not depend on his private information ($d \leq -2$ or $d \geq 2$). Once it happens, further accumulation of public information stops, and conformity of the

[3]BHW assume that when the expected payoffs are equal, the agent randomises (tosses a coin). Although probabilities to be locked in a cascade differ slightly for the different tie-breaking rules, the results are essentially the same.

actions arises. Indeed, after the first agent who decides to discard his private signal and 'join the herd', each of the following agents will face exactly the same decision problem and should 'join the herd' as well. If the system converges to adopting the technology it is said to be an *UP cascade*, if, in contrast, it converges to rejecting the technology it is said to be a *DOWN cascade*. Without loss of generality in what follows we will assume that the **true value of the technology is 1**. Therefore an UP cascade is the 'correct' cascade.

With the signal probabilities and the decision rule one can find some characteristics/statistics of the process which will be useful for the analysis of the informational cascades in a network. These characteristics include average payoff ($W \in [0, 1]$), the share of the agents who end up in an UP cascade, S_{up}, and the share of unlucky ones in DOWN cascades, S_{down}. One can also interpret S_{up} and S_{down} as the probabilities for an agent to find himself locked into an UP or a DOWN cascade respectively. The expressions for W, S_{up}, and S_{down} can be found in Appendix A (A.1-A.3).

3.3 Informational cascades in a network

The BHW model of informational cascades offers a compelling explanation of imitation and herding behaviour observed in many circumstances (see (Bikhchandani, Hirshleifer, and Welch 1998) for a review) based on individual rationality. Not the least, as already has been mentioned above, the predictions of the model are in good terms with the results of experiments. Nevertheless, we cannot use the model for our purposes without some adaptation. Large-scale social or economic phenomena such as diffusion of a new technology or social norms occur in large populations which are often scattered in space. Therefore to model such a process one would have to take into consideration that actors hardly have an opportunity to observe actions of all other agents in the population. The scope of such information is rather limited to the actions of their close neighbours with whom they interact/communicate on a regular basis. As a result the structure of the interactions/communications might come into play.

There are two ways in which ICTs may affect the structure of communication in the society. The first effect is known as the 'death of distance', i.e. decreasing importance of geographical proximity for making and maintaining a contact. The other effect is the increasing density of the communication network, or put it differently, the growing size of information-sharing communities. Although the ICT revolution is likely to increase both the geographical scope and the density of the communication network, these two effects might have different implications for the social welfare. Thus our model should be formulated in such a way that it would allow us examine the effects separately.

3.3.1 The model

To model informational cascades in a structured population we shall modify the basic BHW model described above. Instead of assuming that agents observe the whole population of the size N, we shall assume that they can observe only a subset of the agents with whom they communicate on the regular basis. The communication structure of the population can be viewed as a graph G: a node i of G represents agent i, a (non-directed) edge ij is the 'informational channel' connecting i and j. The degree of the node i, k_i, is the size of the ith neighbourhood.[4]

Following Bala and Goyal (1998) we assume that the rationality of our agents is bounded: an agent does not attempt to infer the decisions of unobserved agents (neighbours of neighbours and so on) from the actions of observed agents (neighbours). One may think that our agents are ignorant of the world outside their local communities. As for the agents' perception of communication structure within their neighbourhoods we assume that an agent believes that none of his neighbours are neighbours of each other, and therefore he treats observed actions as independent (conditional on V).[5]

Because the range of all possible network configurations for any interesting number of agents, N, and edges, K, can be extremely large, we limit ourselves to the study of informational cascades in a particular class of networks - networks generated by 'β-algorithm' used by Watts (1999). According to this algorithm a network is constructed from a lattice graph (1-lattice), where each node is connected with k of its closest neighbours, by rewiring each of its edges with some probability β. Therefore a network is characterized by only three parameters: the number of nodes (N), average degree of a node (k), and the rewiring probability (β). An advantage of this approach is that while random (and therefore generic), 'β-networks' have similar topology (in terms of the path lengths, clustering and so on (Watts 1999)) for the same parameter β. The algorithm allows us to generate a wide spectrum of networks with the given degree of randomness ranging from perfect lattice at $\beta = 0$ to a random graph at $\beta = 1$ for the same number of edges, $kN/2$ (Figure 3.1).[6] Moreover, our experiments with some other types of networks (and with substrates other than 1-lattice) suggest that the results reported here are rather general.

With the 'β-algorithm' we can disentangle the two effects of the globalization of information flows mentioned above. Indeed, since parameter β relates to the share of the distant links, the 'death of distance' effect can be modelled by increasing β.

[4]The communication structure of the BHW model would be represented by a complete graph of N nodes.

[5]Although it is quite a deviation from the original setting of BHW model, it would not change the agents' decision rule.

[6]Our simulations show no difference between networks with $\beta = 1$ and random graphs with the same number of edges ($kN/2$). Therefore in what follows we will use the terms interchangeably.

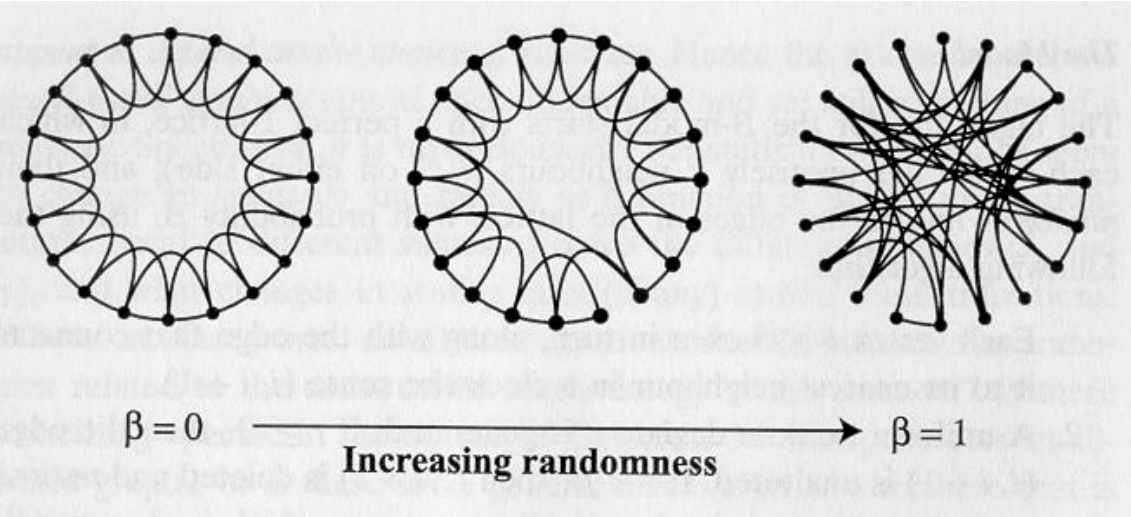

Figure 3.1: Watts' 'β-networks'. For $\beta = 0$, the original 1-lattice is unchanged; for $\beta = 1$, all egdes are rewired randomly, and for $0 < \beta < 0$, graphs combining elements of order and randomness are generated. Adapted from Watts(1999).

Furthemore, for a given number of agents the average degree of a node, k, gives us the average size of information-sharing community, and we can analyse the effect of the density of the communication network on the social welfare by changing k (for fixed N and β).

3.3.2 Results

Figure 3.2 presents the results of our simulations of informational cascades ($p = 0.75$) in a range of networks with $N = 200$, for $\beta = 0$ (1-ring), and $\beta = 1$ (random graph). For each value of k and β we generated 500 networks, and for each of the networks we drew 50 realizations of signal-agent sequences (signals and the order of their arrival). Values of S_a, S_{up}, S_{down}, and $S_{cascade} \equiv S_{up} + S_{down}$ were estimated for each realization and averages reported.

Two approximations to our model for which we can easily find exact solution can be instrumental for understanding the results of our simulations. The first is the BHW model discussed above, with n equal to the size of average neighbourhood of G, which is $(k + 1)$. For $k << N$ one may think of it as an approximation of G by 'caveman' graph, a graph which consists of $N/(k + 1)$ isolated subgraphs ('caves'), where each subgraph is a complete graph of $(k+1)$ nodes. Characteristics S_a, S_{up}, S_{down} as functions of k are given by corresponding expressions for the BHW model (Appendix A, $n = k + 1$) and shown at Figure 3.2 (red circles). At the limit of $k = N - 1$ (complete graph) our simulations for informational cascades in all

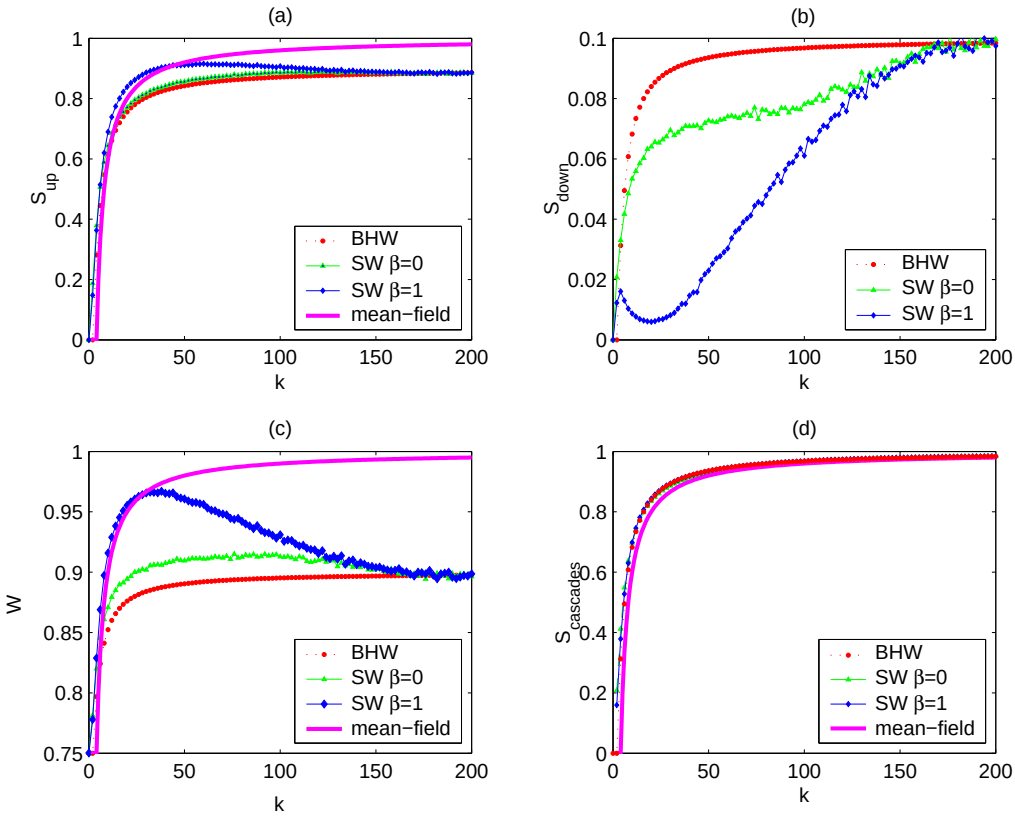

Figure 3.2: Effect of the network density, the number of edges $kN/2$ ($N = 200$, $p = 0.75$): (a) Share of agents in UP cascade, S_{up}; (b) Share of agents in DOWN cascade, S_{down}; (c) Social welfare (total payoff normalized by the size of papulation), W; (d) Share of agents locked in cascades, $S_{cascades}$. β-networks: random graph ($\beta = 1$)- diamonds, 1-lattice/ring ($\beta = 0$) - triangle. The BHW/'caveman'model - circles. Mean-field solution - solid line (there are no DOWN cascades in mean-field solution (see Appendix B), i.e. $S_{down} = 0$.)

networks must produce the same results as BHW with $n = N$.

The other approximation uses the 'mean-field approach' (Weisbuch, Kirman, and Herreiner 2000). In this approach we approximate actual realizations of random variables by corresponding values averaged over the population and solve dynamics for those average values. Hence the share of positive signals is approximated with the probability of the positive signal, and we would examine the behaviour of a 'representative agent' in a 'representative neighbourhood' (Appendix B). In doing so we neglect possible fluctuations of the random variables assuming that in the limit of large population the fluctuations may have only minor impact on the dynamics of the system. The mean-field solution for our model is shown by dotted line at Figure 3.2.

As one could have already guessed, these two approximations lie at the extremes

from the point of view of the strength of local effects. In the 'caveman' world local cohesion is at maximum: within the local community all agents are neighbours of each other. As a result there is high correlation between neighbours' choices. In contrast, the mean-field approach denies possible local fluctuations,[7] thus there is no relationship between neighbours' choices other than on the level of the whole population.

Note, that even though all networks generated by β-algorithm are connected in the graph-theoretic sense (Watts 1999), i.e. for any pair of nodes i and j there is a path connecting them, for a small k the society is likely to be "effectively disconnected": for the action of agent i to affect the choice of agent j, by the time when agent j is to make his decision all agents on one of the paths connecting i and j must have made their choice, which is rather unlikely when k is small (average distance in G is large). For this reason interactions effectively 'localized' within a community.

These local effects do make a difference. Due to informational externalities the model has a self-reinforcing ('positive feedback') mechanism, and sufficiently strong fluctuations have potential to grow up to the size of the local community. But the strength of those effects depends on the structure of the network. In a network with high intensity of local interactions we may expect that the local effects prevail and cascades will be formed in each of the local communities almost independently, therefore BHW provides a reasonable approximation; while in a network with loose structure of local contacts, local fluctuations do not have significant impact on the overall outcome, and the process will be governed by averages over the population, hence a mean-field approach does better in this case.

It seems that the results of our simulations are consistent with the logic explained above (see Figure 3.2). Indeed, the 1-lattice ($\beta = 0$) obviously has much more prominent local structure than a random graph ($\beta = 1$), and as one can see the results for 1-lattice can be fairly well approximated by the BHW/'cavemen' model. The results for the random graph, on the other hand, go particularly well with the mean-field solution for $k << N$. As k increases and comes close to N, the global local effects starts to dominate the picture (the size of the local community approaches the size of the whole population), therefore the results for all networks expectedly converge to the corresponding values of the BHW model ($n = N$).

3.4 Discussion

From the point of view of the social planner it is essential to know which kind of the communication structure (represented by our graphs) does better in terms of the

[7] both fluctuations of signal-agent series and of structure of the generated network, G.

social welfare. In our model social welfare could be approximated by overall social outcome, the sum of agents' payoffs normalized by the size of population.

A remarkable feature of Figure 3.2 is that for the same number of edges a random graph significantly outperforms 1-lattice (except the region of very small k, where threshold effects cannot be neglected). To explore this point more thoroughly we run simulations fixing k and changing the rewiring parameter β. The results of the simulations are shown in Figure 3.3 ($N = 200$, $k = 10$). As one can see 'randomisation' of the structure by rewiring edges, which creates more of 'distant' links at the expense of the 'local' ones, increases the social welfare: the overall outcome grows fast until it reaches plateau at $\beta \sim 0.5$. This increase seems to be related to the opposite effect that rewiring has on the shares of UP and DOWN cascades (Figure 3.3a and 3.3b). The origin of this effect is rather clear. In a network with high local cohesion, such as 1-lattice, cascades emerge almost independently in each of the local communities and we would see a set of subpopulations some of which are evolving toward UP cascades and others toward DOWN cascades.

For large N/k the share of subpopulations in UP cascades must be about the probability of UP cascade in BHW model, and the same for DOWN cascades. If we start to create links between such local communities the agents who get links outside their subpopulation become exposed to 'practices' in other communities. Because the share of communities in UP cascades is much higher then the share of agents in DOWN cascades, an agent with 'external' links is more likely to be connected to the communities in UP cascade. This would not cardinally change the agent's perception of the technology, if the agent already belonged to the community which is going to converge to UP cascade, it would only reinforce the tendency. Consequently, the share of agents locked in UP cascades increases with rewiring. In contrast, if the agent belonged to a community which tends to converge to a DOWN cascade rewiring would have opposite effect, damping or even preventing the growth of the 'negative' fluctuation. One can also see it from Figure 3.2 (diagrams 3.2a and 3.2b).

Another striking feature of Figure 3.2 is that an increase in the density of connections (increase in k) does not always improve social welfare. Indeed, for a network "more connected" than caveman graph the social outcome W is not a monotonic function of k: first, it rapidly increases with k until some k^*, and then starts to fall. The explanation is straightforward. In a dense network ($k \sim N$) the size of local communities is comparable with the size of the population. As we already know, small 'negative' fluctuations may spread through the community, which now has size $\sim N$, leading to the outcome suboptimal with respect to social welfare.

Do our results differ from those of BG? To see that let us first briefly restate the results of their work relevant to our study. BG consider 'boundedly rational' agents repeatedly experimenting with technologies producing stochastic outcomes. The expected outcome of a technology is determined by the state of nature. Agents employ Bayesian learning in updating their beliefs about the state of the nature (the quality

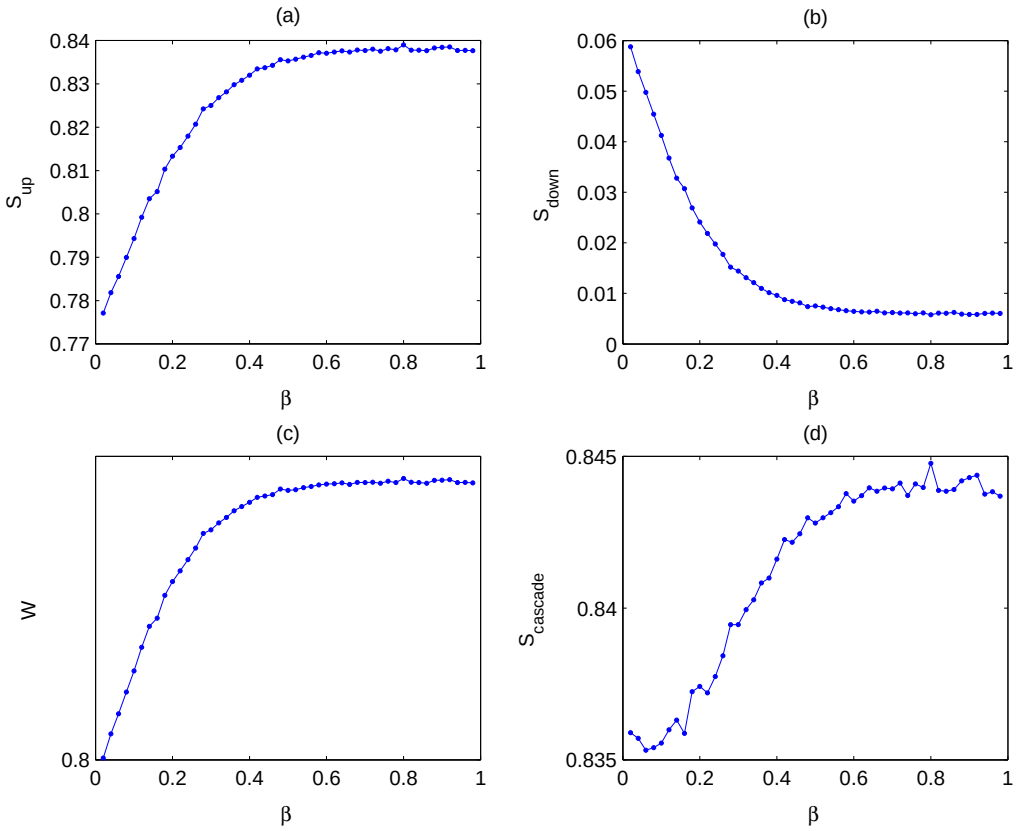

Figure 3.3: Effects of rewiring ($N = 200$, $k = 10$, $p = 0.75$). (a) Share of agents in UP cascade; (b) Share of agents in DOWN cascade; (c) Social welfare (total payoff normalized by the size of papulation); (d) Share of agents locked in cascades. Randomness increases with β.

of the technology) with the experience of their own and the other agents with whom they have links (neighbours). In the settings of (Bala and Goyal 1998) (more general than (Bala and Goyal 2001)) BG proved that in a connected society of homogeneous agents the better technology always drives out inferior ones (complete social learning), if the following conditions on beliefs and the structure of the relationships are satisfied: (a) agents' priors are dispersed enough (else agents may be stuck with the inferior technology from the beginning)[8]; (b) there is no 'royal family' (or 'nearly royal families'), which is a finite subset of the population such that all agents from this subset are 'visible' to all other agents in a network, or the size of the 'royal family' is sufficiently small with respect to the size of agents' neighbourhood (to ensure that a 'bad' experience of the 'royal family' will not overwrite locally emerging knowledge). In addition, in the case of agents with heterogeneous preferences lock-in to an inferior technology may happen if the society is not 'group-wise' connected,

[8] Condition H in (Bala and Goyal 2001).

i.e. when an agent (or several agents) of one type is surrounded by the agents of the other type effectively separating this agent from others of his type (Bala and Goyal 2001).

Let us hypothesize what may happen in the world of BG agents, if we start to increase the number of informational channels (add edges into G). First, and obvious, at some point a disconnected society (such as represented by a 'caveman graph') will become connected, that increases the chances that the better technology will prevail. Second, one may think that the heterogeneity of priors is 'effectively' increasing as we connect previously disconnected groups, hence decreasing the probability of an 'inferior lock-in'. Third, if we add more new links, the effect of the 'royal family' (if one existed in the initial network) would be weakened because each agent's 'reference group' is growing. And finally, increasing the density of the communication network in the society of heterogeneous agents would lead to higher chances of 'group-wise' connectivity.

Although we cannot compare different models, it is almost obvious that implication of the models for the issue we are interested in might be different. Intensification of information exchange expressed in the increasing density of the network links would unambiguously benefit the social welfare in the BG model, while in our model with the BHW model of agents' behaviour it may lead to a 'global lock-in' to inferior practices (although at the beginning of the process, i.e. with small k, it significantly improves social well-being, see Figure 3.2c).

It is not a secret for the reader, that the divergence of our results from the results of BG could have been predicted already from the differences in the settings of their model and the BHW model used by us (the order of decisions, the scope of publicly available information and so on). Hence a would-be productive discussion should shift to the discussion of the realism of the assumptions of the BHW model. However, for such discussion one is better to refer to the original work of BHW (Bikhchandani, Hirshleifer, and Welch 1992, Bikhchandani, Hirshleifer, and Welch 1998)[9]. Here we would limit ourselves only to several remarks on the difference between the assumptions of the BG model and the BHW settings.

First, in many circumstances private information is unlikely to be shared, either because it has commercial value, and transactions costs of buying information from scattered sources might be high, or simply because of its 'tacit' nature. Furthermore not everybody "practices what he preaches", and not every piece of information which one is able to collect can be trusted ("actions speak louder than words"). Second, the payoffs to adopting might be realized only in a distant future (in our model, only after the whole process is already over). We could recall an example of tobacco, which was introduced into Europe, and spread as a cure for headache. Third, not

[9] One can also find the annotated bibliography on informational cascades/herding on the website devoted to informational cascades http://welch.som.yale.edu/cascades/.

always decisions about adoption of certain technology are simultaneous as in BG model (instead of *sequential* order of the BHW model). In addition, switching from one technology to another may be costly, as a result once one has invested in a technology he could be reluctant to replace it with another technology any soon. Should one argue that in 'true' long-run the true state of nature (quality of the technology) will be revealed (as in the example with tobacco) due to some information leakages, and switching costs are spread through the time and not so relevant, we shall remind that, first, in the long-run 'we are all dead'- by the time the society should figure out the true quality of the technology, this technology may be substituted by another (perhaps inferior as well). Consider also that the consequences of today's decisions may turn to be irreversible.[10] Not the least to mention, humans tend to rationalize their choices ex-post, and, arguing over today's practices, which have 'survived' in the evolution, one would have to resort to counterfactual reasoning not very welcomed in the modern economic theory.[11]

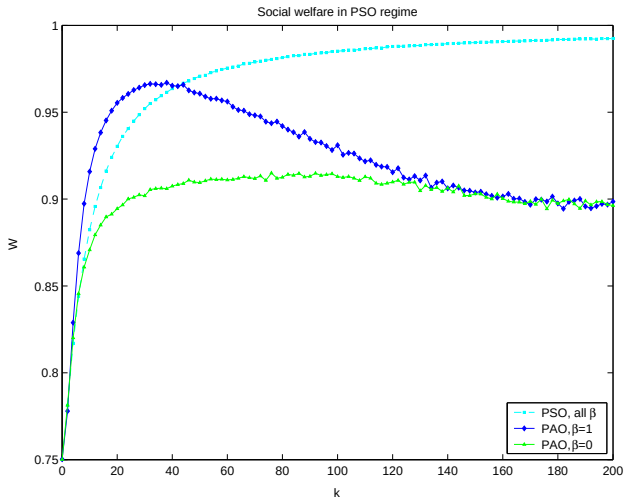

Figure 3.4: Social welfare as a function of network density under previous-signal-observable (PSO) regime, and under previous-action-observable (PAO) regime. Under PSO the results for networks with all β are the same.

It seems that the assumption most crucial for our results concerns with the scope of information available to an outside observer. In the 'previous-signal-observable' (PSO) regime of the BHW model, where not only actions chosen by agents, but also their signals are publicly observable, the true state of nature will be revealed relatively fast therefore socially optimal outcome will be reached, and no cascade

[10]Not too much we can do about global warming today, which is likely to be related to air pollution in the last century. Even less could be done about some animal species which disappeared from the surface of the Earth once and forever, directly or indirectly due to human activities.

[11]see (Cowan and Foray 2002) for a discussion.

will emerge. We may expect similar results in a network. In addition, increasing density of the communication network is likely to speed up the process of learning. Indeed, the results of simulations for PSO regime in our model presented in Figure 3.4 show that social welfare monotonicly increasing with the density of the network.

3.5 Conclusions

Now we are ready to answer the question raised in the beginning of this paper: is globalisation of information flows always good for human society? From (fairly limited) prospective of our model we can say that all depends on how this process affects the structure of communications in our society. Namely, if the ICT revolution is about making the Big World small through enhancing exchange information between distant communities (which corresponds to rewiring in terms of our model), the answer would be '*yes*'. Exposing ourselves to experiences from different parts of the world might prevent a lock-in to inferior practices. In this point we share the conclusions of BG. In addition, our model suggests that we can expect the effect of such globalisation to be higher at the beginning (when β is relatively small) than at the end of the process (interim as well) as one can see from Figure 3.3.[12] This is the 'bright side' of globalisation.

On the other hand, decreasing costs of communication and bringing new technologies for data handling[13] improvements in ICT increase not only the geographical reach of information flows, but also the size of 'information sharing' communities which makes the whole world one local community (k increases and approaches to N), which would lead to convergence to a uniform culture (perceptions, norms, values etc.) and elimination of diversity. This should not be cause for concern, if the 'global culture' were to select only the best practices from the variety of those existing today, eliminating inferior ones. The bad news is that there is no such a warranty. It might well happen that the historical (path-dependent) process would lead us to an inferior state of affairs in so far as social well-being is concerned.

There are several ways to deal with the 'global informational lock-in problem' which follow directly from the model (if we exclude radical, but unfeasible solutions such as restrictions of public access to ICT). The first one coming from the studies of technological lock-ins is to prevent 'early standardization' (early emergence of informational cascades) and encourage diversity of opinions, for instance support views alternative to the mainstream. Another recipe is to promote wide discussion and engage different communities into dialog, so to stimulate agents to reveal their private information (PSO regime of BHW). As we have seen under PSO regime as

[12] Compare with BG's 'degree of integration' η.

[13] And institutional changes related to it (Petersen and Rajan 2002, e.g.,).

in the model of BG increasing density of information flows unambiguously enhances social welfare.

Chapter 4

Mutual Illusions and Financing New Technologies: Two-Sided Informational Cascades

4.1 Introduction

Many authors writing on human behavior have noticed that individual decisions are often influenced by decisions made by others. They have documented a number of situations in which individuals prefer to follow the 'crowd', while their own feelings are against it. There is a range of different social mechanisms which may cause conformist behavior of individuals such as punishment of deviators (Akerlof 1980), positive payoff externalities (Arthur 1994) and so on. It is also possible that herding arises as consequence of the bounded rationality of individuals (Shiller 1989).

In the last decade there has been a surge of interest in a particular kind of mechanism behind conformist behavior, which can explain *voluntary rational* 'herding'. After the seminal works of Bikhchandani, Hirshleifer, and Welch (1992) and of Banerjee (1992) this social phenomenon is often referred to as an 'informational cascade'.

The basic idea of informational cascades is that in certain environments where private information can be revealed only through individual actions because of information externalities, truly rational agents may find it optimal to follow the choice of others, rejecting their own information. Perhaps the most striking feature of informational cascades is that when there is noise in private information there is always a positive probability that the overall outcome will be suboptimal, i.e. agents will form a cascade in which they reject optimal actions in favor of inferior ones regardless of their private information. Therefore, information structures that are vulnerable to information cascades in this way can have negative effects on social

welfare.

The original BHW model has been extended in a number of ways and the robustness of the model with respect to changes in assumptions has been examined. The informational cascade framework has been used to explain a wide range of social phenomena such as fads, fashions, medical (mal) practice, collapse of political regimes among the others (see (Bikhchandani, Hirshleifer, and Welch 1998) for a review). There are also some important applications of this kind of model in economics and finance. Welch (1992) applied informational cascades to the IPO market and explains underpricing incentives of issuers. Avery and Zemsky (Avery and Zemsky 1998) argued that short-run mispricing on financial markets may be a consequence of investors' herd behavior.

Most of the informational cascade models assume that the agents forming a cascade are the same either with respect to the sort of information available to them, or with respect to roles they play, or both. For instance, in Avery and Zemsky's model of financial market investors have different roles: some of them are sellers and others are buyers of assets, but the sort of information available to agents is essentially the same.

However, there is no *a priori* reason to suppose that the agents on the two sides of the market have identical information sets, and modelling some situations may require us to take into account that two sides of the market having access to different information.

Consider financing new technology. It is widely recognized today that one of the main problems of external financing in new high-tech industries is information asymmetry between firms developing new technologies and their potential investors (Branscomb and Auerswald 2001). On the one hand, financial institutions and individual investors often do not have enough expertise to judge 'state-of-the-art' technologies. On the other hand, firms working in new industries, especially new small ventures which mainly contribute to development of 'at-the-edge' technologies, have problems with evaluating both market and financial potential for the products they are developing. It seems quite reasonable to assume that investors have better knowledge of the *market perspectives* for new technologies, while firms know the *technology* with which they are working. This is an example of the situation in which information sets available to the opposite sides of the market are different.

There is also no *a priori* reason to believe that only one side of the market is subject to information cascades. In our example of financing new technology the process of acquiring information about the technology and about the market for new products is costly, and therefore has commercial value. As a consequence, agents have incentives not to share their private information and so spillovers of this kind of knowledge are limited, at least in the short term. Nevertheless, their actions, or 'outcomes' arising from the actions, in many cases are observable and it may create conditions necessary for informational cascades. This argument must be valid for

both entrepreneurs and venture capitalists. Hence we may expect informational cascades on the two sides.

In this paper we examine a simple setting of two-sided informational cascades and show that taking into account both sides of the market with different information sets may generate interesting learning dynamics on both sides of the market. Like in the model of Bikhchandani et al. the scope of public information is likely to affect the social welfare. We also find that when the information asymmerty is high, both sides of the market tend to be 'overoptimistic'.

4.2 Financing New Technologies

Innovation management literature distinguishes two main sources of the risk in the process of development and commercialization of a new technology. *Technical risk* is the risk of "failure in the attempt to convert invention to innovation (as when the product does not meet specifications in terms of performance, cost of production, or reliability)." *Market risk* arises due to the uncertainties about the demand for the product that might not be formed yet, and may include the risk " that the product will [not] provide vectors of differentiation sufficient to distinguish it from competitive offerings" and the risk "that the proposed business model will [not] be successful in the market"((Branscomb and Auerswald 2001), p.4).

One may visualize these risks with the "quadrants of risk" diagram reproduced in Figure 4.1 (Hartmann and Myers 2001). A technological venture has the lowest risk when the product uses the technology which is relatively well known, and the market for the product already exists. This is the case of 'evolutionary' innovation. In contrast to the evolutionary innovation a 'radical' one employs a state-of-art technology, and the product is to be brought to the new markets. Although a 'radical' project has the highest risk, in case of success it is more likely to offer the great opportunities for future growth.

While large enterprises have natural advantage operating in established markets, they are often too slow to react to the emerging opportunities offered by radical innovations and new markets. This opens a room for small technological ventures. As an inside observer put it in the IT sector "perhaps half of the industry growth over five-year period . . . will be captured by newly , previously unrecognized players, offering new products and services, building on new business models" (McGroddy 2001).

However unlike their larger counterparts, small ventures rarely can finance such projects themselves, and thus have to rely on external investors. One of the major problems of external financing is the 'information and trust gap':

> "On each side of the Valley of Death stands a quite different archetypal character: the technologist on the one side, and the investor/manager on the other.

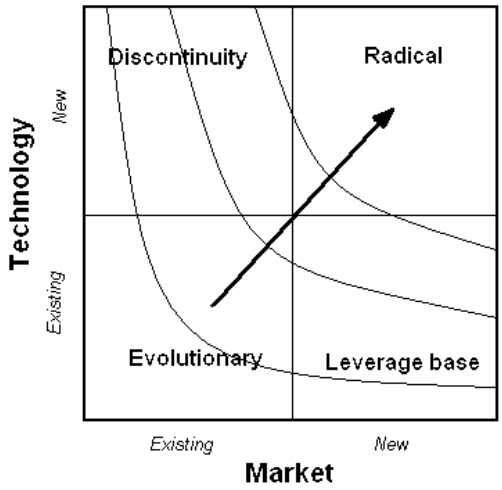

Figure 4.1: Quadrants of risk. Adopted from Branscomb and Auerswald (2001).

> Each has different training, expectations, information sources, and modes of expression. The technologist knows what is scientifically interesting, what is technically feasible, and what is fundamentally novel in the approach proposed. [...] The investor/manager knows about the process of bringing new products to market, but will likely to have to trust the technologist when it comes to technical particulars of the project in question.[...] To the extent that technologist and investor/manager do not fully trust one another or cannot communicate effectively, the Valley of Death between invention and innovation becomes deeper still." (p.12 (Branscomb and Auerswald 2001))

The 'information and trust gap' has received significant attention in the recent years because of the important role which small enterprises play in rapidly growing high-tech industries. In particular, much research effort has been focused on understanding how the problem of asymmetric information affects the mode of financing (multistage investment, VCs' syndication, etc.), exit strategies, venture capital structure etc. (see (Gompers and Lerner 1999) for a review). However, so far the unit of analysis in this field has been single VC-firm relationship, and we are yet to study intra-population effects (e.g. interdependencies between the VCs' choices in terms of technologies).

In the next section we introduce a model of the process of social learning within the populations of entrepreneurs and investors in the presence of the 'information and trust gap'. The backbone of our model is the model of informational cascades of Bikhchandani, Hirshleifer, and Welch (1992) (BHW henceforth). However to take into account the effects of two-side interactions we modify the original model in

several ways. First, we assume that success of the venture is depends on two factors: the state (potential) of the technology and market prospects of the future product. Second, the information on the states of the technology and market is distributed asymmetrically: entrepreneurs know the technology, while investors have better knowledge of the market. Third, we consider a situation where public knowledge consists of the outcomes of negotiations rather than actors' actions. We examine how the probabilities of two-sided cascades depend on the degree of uncertainty about technology and market, as well as on the scope of public information.

4.3 The model

Actors There are two populations: potential investors, who, for convenience, we call venture capitalists (VCs) and the population of entrepreneurs.

States of the World Success of a project is determined by two factors: by technology itself (e.g. quality of the product if the project involves product innovation, productivity gains if it is about process innovation), and by the market prospects for the new technology (e.g. its prospective demand). As in BHW we assume the state of technology, E, and the state of market, V, be binary variables: $E \in \{h, l\}$, $V \in \{H, L\}$, with equal prior probabilities of 1/2.

Payoffs There is a market place, where an entrepreneur meets with a venture capitalist to discuss the potential project. If any of the sides chooses to *Decline* the deal, the negotiations fail, no project takes place, and both sides stay with their reservation values, which without loss of generality are set to be zero.

If the negotiations have been successful, and project starts, then the payoffs to the contracting sides are determined by the state of the world. If both technology and market are high ($E = h$, $V = H$), the project will have success, and both parties will gain from the project. When one of the sides, say market, is low, but the technology has a great potential ($E = h$, $V = L$), then the project may still be successful. To achieve this success the entrepreneur will have to work hard, and his payoff in this case is below his reservation value, while the investor in this case will be a 'free-rider', and his payoff exceeds the reservation value. Similarly, if the technology is mediocre, but the market is high ($E = h, V = L$), the venture capitalist will have to put in more effort to ensure the success. In the case of both sides being mediocre ($E = l$, $V = L$), both parties have payoffs below their reservation values. We also assume that once the venture started, it is not in the interest of the agents to disrupt it. [1]

[1] For example, that might be the case if an immediate liquidation of the project would severely

Table 4.1 summarizes the payoffs. The values are chosen so that from the both sides payoffs are the same as in the BHW model.

Table 4.1: Payoffs (EP, VC)

	$V = H$	$V = L$
$E = h$	(1, 1)	(-1, 1)
$E = l$	(1,-1)	(-1,-1)

The agents choose their actions maximizing expected payoffs, which are based on public information and their private signals. In the case of a 'draw' i.e. when the expected payoffs from both actions are the same, we assume that agents trust more to their own intuition, and take the decision according to their private signals.[2]

Information structure We assume that entrepreneurs observe the state of technology E, while venture capitalists observe the market prospects for the new technology V; but neither do entrepreneurs know V, nor do venture capitalists know E. That is, entrepreneurs and venture capitalists have different information sets.

We model agents' subjective 'guesses' about the state of the other side of the market as private signals of limited precision. The t-th entrepreneur observes a conditionally independent identically distributed signal $v \in \{H, L\}$ about state of V, and the t-th venture capitalist gets a signal $e \in \{h, l\}$ about E. Tables 4.2 and 4.3 describe the signal probabilities ($p, q > 1/2$).

Table 4.2: Signal probabilities for entrepreneurs

	$\Pr(v = H\|V)$	$\Pr(v = L\|V)$
$V = H$	p	$1 - p$
$V = L$	$1 - p$	p

Each period t=1,2,... a pair of agents meets. The project goes ahead only if both sides agree to participate. Thus, the outcome of the negotiations is *Proceed*, or *Not Proceed*.

damage their reputation, and in this way the cost of liquidation exceeds their losses from continuation with the project.

[2]This kind of tie-breaking rule was employed by Anderson and Holt (1997) in their experimental study of informational cascades.

Table 4.3: Signal probabilities for venture capitalists

	$\Pr(e = h\|E)$	$\Pr(e = l\|E)$
$E = h$	q	$1 - q$
$E = l$	$1 - q$	q

We consider two information structures that differ in the scope of the public information about the past. In previous-action-observable (POA) model, as in BHW, the information about **actions** (*Agree* or *Decline*) chosen by all agents in the past is available to public, while in previous-outcome-observable (POO) model only information about **outcomes** (*Proceed* or *Not Proceed*) of the negotiations becomes public.

As one might expect, and it will be shown in the next section, POA model is essentially the same as BHW. Therefore, comparison of the two models may provide us an idea of how the limitations of the public history affect the probability of 'incorrect herding'.

4.4 Analysis

Our analysis proceeds as follows. First, we examine a one period game with exogenously-given agents' beliefs. Then we will turn to how the beliefs in the multiperiod setting are formed under Bayesian learning in POO model. The section concludes with the general set up for the multiperiod POO model.

However, before we start to examine POO model, we will discuss POA model which is almost the same as the (POA) model in BHW. Later, when we will discuss the results of our simulations for the POO model, the POA model will be used as a benchmark.

4.4.1 Model with observable actions

The analyis for the 1-period game that will be developed in the next section can be applied to the model with publicly observable actions. However, analysis of POA can be made with much more simpler means by analogy with BHW.

Notice that the payoff to an agent is entirely determined by the state of the other side, and does not depend on the state of the world on his side, as seen in the payoff matrix, Table 4.1. Entrepreneurs only care about V, and venture capitalists are interested only in E. It follows that once an agent, say an entrepreneur, has an opportunity to observe actions of the entrepreneurs that are driven by their

feelings about V, and the actions chosen by the other side has no value for him since venture capitalists' actions depends on venture capitalists beliefs about E which is of no interest for an entrepreneur. Due to this in POA model we have two sides that lock into one-sided cascades independently.

The only thing we have to be cautious about is that a cascade in actions does not necessarily mean a cascade in outcomes, which is our primary interest. While a DOWN cascade on any side (as it is defined in BHW) always results in an infinite negative series in the outcomes, an UP cascade on one side may not result in a never ending series of positive outcomes yet, since the other side may be declining the offers. Therefore, we can say that a DOWN cascade in two-sided POA model happens when either or both of the sides rejects the deals regardless of signals. Two-sided UP cascade happens when *both* sides agree whatever is their private information.

One can easily find that as in BHW one of the sides starts a cascade when the agents on this side receive two (or three) similar signals in row.[3] If signals are negative a DOWN cascade emerges, if they are positive an UP cascade arises. Two signals of different signs cancel each out, and the following agent finds himself in the same situation as the agent two periods before him, e.g. if $v_0 = H$, $v_1 = L$, then the third entrepreneur has the same prior belief about V as the first one. The probability for entrepreneurs to be locked in UP cascade is

$$\Pr(EP\ in\ UP\ cascade) = \frac{p_V^2}{1 - 2p_V + 2p_V^2}.$$

where $p_V = \Pr(v = H|V)$ (Table 4.2).[4] Similarly for venture capitalists' UP cascade

$$\Pr(VC\ in\ UP\ cascade) = \frac{q_E^2}{1 - 2q_E + 2q_E^2}.$$

where $q_E = \Pr(e = h|E)$ (Table 4.3).[5]

Since a two-sided UP cascade is nothing more than two simulataneous independent UP cascades on each side, the probibility to end up in two-sided UP cascade is

$$\Pr(two\text{-}sided\ UP\ cascade) = \frac{p_V^2 q_E^2}{(1 - 2q_E + 2q_E^2)(1 - 2p_V + 2p_V^2)}. \tag{4.1}$$

since the system in BHW model converges to one of the cascades with probability

[3]This is given the tie-breaking rule we use. BHW use different tie-breaking rule and have somewhat different results including the formula for the probability of a cascade.

[4]Consider limit of the right-hand side of the equation A.1 at $K \to \infty$. Notice the difference in the notations ($p = p_V$, $q = 2p_V(p_V - 1)$).

[5]In the notations of this model: $p = q_E$, $q = 2q_E(q_E - 1)$.

1, the probability of two-sided DOWN cascade is

$$\Pr(\textit{two-sided DOWN cascade}) = 1 - \frac{p_V^2 q_E^2}{(1 - 2q_E + 2q_E^2)(1 - 2p_V + 2p_V^2)}. \tag{4.2}$$

The ease with which we manage to examine the POA model comes from the payoff matrix that we choose to be similar to payoffs in BHW model, and from the fact that before an information cascade arises agents can infer (from the actions) the private signals their predecessors received.

The latter is not the case in the POO model. From a negative outcome an outside observer cannot infer actions and therefore the private signals that the pair received.

4.4.2 Model with observable outcomes

One-period game

Consider a pair of entrepreneur and venture capitalist who are to decide upon setting up a project. For the moment we assume that their beliefs (based on their private signals and the public history of previous negotiations) are exogenously given.

We use a static Bayesian game with 4 types of players on each side to analyse agents' decisions. The type of a player in this game is determined by the state of the market on his side (E for entrepreneurs, V for venture capitalists), and the private signal he receives (v or e).

Let P_{Ev} be the probability that the market is in a high state ($V = H$), given that the state of the technology is E, and he has received signal v (i.e. his type is Ev). Similarly, Q_{Ve} is venture capitalist's belief that $E = h$, when the state of the market for the new technology is V and the private signal about technology is e. We require rather natural conditions on the beliefs: $P_{EH} \geq P_{EL}$, $Q_{Vh} \geq Q_{Vl}$, i.e. a positive signal strengthens (at least does not weaken) one's belief that the other side is in the favourable state. The probability for a venture capitalist to get positive signal conditional on that the state of the technology is E will be denoted as q_E, and the conditional probability for the entrepreneur to receive a positive signal is p_V.

The signals are independent, hence, for example, the belief of an entrepreneur of type Ev that he is playing with the venture capitalist of type Lh is $(1 - P_{Ev})q_E$. A venture capitalist's belief that his opponent belongs to the type hL is $Q_{Ve}(1 - p_V)$, given that VC's type is Ve and so on.

Each of player has a choice between two actions (*Agree* or *Decline*). The normal form of the game is presented in Table 4.4 (letters A and D stand for players' actions: *Agree* or *Decline*).

Table 4.4: 1-period game in the normal form

(EP, VC)			P_{Ev}				$1-P_{Ev}$			
			$q_E P_{Ev}$ ***Hh***		$(1-q_E)P_{Ev}$ ***Hl***		$q_E(1-P_{Ev})$ ***Lh***		$(1-q_E)(1-P_{Ev})$ ***Ll***	
			A	***D***	***A***	***D***	***A***	***D***	***A***	***D***
Q_{Ve}	$p_V Q_{Ve}$ ***hH***	***A***	(1, 1)	0	(1, 1)	0	(-1,1)	0	(-1, 1)	0
		D	0	0	0	0	0	0	0	0
	$(1-p_V)Q_{Ve}$ ***hL***	***A***	(1, 1)	0	(1, 1)	0	(-1,1)	0	(-1, 1)	0
		D	0	0	0	0	0	0	0	0
$1-Q_{Ve}$	$p_V(1-Q_{Ve})$ ***lH***	***A***	(1,-1)	0	(1,-1)	0	(-1,-1)	0	(-1,-1)	0
		D	0	0	0	0	0	0	0	0
	$(1-p_V)(1-Q_{Ve})$ ***lL***	***A***	(1,-1)	0	(1,-1)	0	(-1,-1)	0	(-1,-1)	0
		D	0	0	0	0	0	0	0	0

Now we will examine what are (Bayes-Nash) equilibria in this game for different sets of beliefs. We will limit our analysis, considering only equilibria in the pure strategies.

Equilibria Player's strategy (decision rule) in a Bayesian game is a set of the actions for all types of the player. There are 16 strategies available for each player in our game. To denote them we will extend notations of Table 4.4.

In this notation the entrepreneur's strategy is a string $a_{hH}a_{hL}a_{lH}a_{lL}$ where $a_{hH} \in \{Agree, Decline\}$ is the action played when an entrepreneur is of type hH, a_{hL} if he is of type hL and so on. Similarly $b_{Hh}b_{Hl}b_{Lh}b_{Ll}$ stands for venture capitalist's strategy such that a venture capitalist chooses action b_{Hh} when he is of type is Hh and so on.

A Bayes-Nash equilibrium (in pure strategies) for a Bayesian game is the strategy profile such that each of the players chooses a best response to the conditional distribution of his opponents' strategies *for each type that he may belong to.*

Table C.1 presents payoffs to each type of entrepreneur. Since an agent gets zero payoff if he *Decline*s, we need only consider payoffs to the action *Agree* when played against different strategies of his opponent. For every type of the entrepreneur if the expected payoff of *Agree* is negative he must play *Decline*, and *Agree* whenever it is positive. In the case of a 'draw' (i.e. zero expected payoff),according to our tie-breaking rule, we assume that one follows his private signal: *Agree* if the signal is favourable, ans *Decline* otherwise.

Similarly, Table C.2 presents VC's expected payoffs if he plays *Agree* against different strategies of his opponent.

Examining of Tables C.1 and C.2 one might note that,

- Strategies $**DA$ and $DA**$, where '$*$' stands either action *Agree* or *Decline*, should not be played in an equilibrium (which is rather natural: why would one *Decline* the deal when he receives a favourable signal, and nevertheless *Agree* if the signal were negative?). Indeed, by our assumption about beliefs: $P_{EH} \geq P_{EL}$ and $Q_{Vh} \geq Q_{Vl}$. Therefore according to Tables C.1 and C.2 the expected payoff of *Agree* to an agent, say, to an entrepreneur of type EH is greater or equal, than to one of type EL. Thus that if an entrepreneur of type EL plays *Agree*, so does the one of type EH. In the case of $P_{EH} = P_{EL}$ or $Q_{Vh} = Q_{Vl}$, the tie-breaking rule applies.

- For any set of beliefs there will be 'status quo' equilibria ($DD**$,$DDDD$) and ($DDDD$,$DD**$). Indeed, on the one hand, if my opponent is playing $DDDD$, I would get a zero payoff whatever strategy I will use. On the other hand, if I am playing $DD**$, then the payoff of my opponent is negative (or zero) whatever is his type, and would therefore be better off refraining from the deal.

- Whenever there is non-zero probability that the other side is in a favourable state, strategy $AAAA$ is the only best response to $A*DD$. Otherwise, one should play $ADAD$ or $ADDD$ in response to $A*DD$.

Note, that although 'status quo' equilibria exist for any set of beliefs, whenever the beliefs are so that there is another equilibrium, hold, 'status quo' equilibria are not *regular*, that is they are very sensitive to small perturbations of the payoffs. For this reason we will not consider those equilibria in the following analisis.

For the similar reason we will ignore equilibria ($ADAD$,$A*DD$) and ($A*DD$,$ADAD$). Those equilibria require that one of the sides has to be sure that the other side of the market is weak (otherwise he should play $AAAA$ instead of $ADAD$). Hence whatever startegy the other side is playing the expected payoff is non-positive. Due to that fact, those equilibria are not stable with respect to 'trembling hand': if the other side is playing $**A*$ or $***A$ instead of $**DD$ with any small but non-zero probability, then the expected payoff is negative and the player should *Decline* regardless to his signal.

Taking into account these remarks there are only 20 possible equilibria in our game that may result in a positive outcome. The conditions on the beliefs necessary for these equilibria are listed in the Table C.3. We will asume that once the conditions for any of these equilibria hold this equilibrium will be played (if there are several possibile equilibria, players will toss a coin). If neither of the conditions hold then an equilibrium with one of the sides playing $DDDD$ strategy will be played (the result would be a negative informational cascade).

So far, we have treated agents' beliefs as exogenously given. However in our model agents form their beliefs from on the history of the outcomes and their private signals. Each period agents observe an outcome of the negotiations, assess which equilibrium has been played, and update their beliefs according to Bayes rule. Appendix C reformulates the model in the 'likelihood ratios' and provides the details of the updating rules.

General set up

As in BHW we start with prior probabilities (P_E^0 and Q_V^0) of 1/2, or in terms of the likelihood ratios $A_0 = B_0 = C_0 = D_0 = 1$. This corresponds to the equilibrium ($ADAD$,$ADAD$), which we can call 'follow-your-signal' equilibrium, since the players choose *Agree* when they receive a positive signal, and *Decline* if the signal is negative.

Once the signals are received they choose their actions according to their strategies. If both players *Agree*, the result will be positive, otherwise it will be negative. The result of the negotiations will be used by the followers to update their beliefs according to the belief update rules listed in Table C.5. Updated beliefs will be used to find the equilirium/equilibria to be played by the next pair etc.

If there are several equilibria that may take place for a given set of beliefs we assume that the equilibrium to be played is determined randomly (each equilibrium has the same chances to be played) and which equilibrium is played will be known to public.

Informational cascades

BHW define an informational cascade as the situation where individual actions does not depend on the private signals. Once it happens the actions of individuals do not reveal any new information to the followers, hence the followers will find themselves in the same situation as their predecessors, therefore, they should ignore their private signals as well. As an informational cascade starts, an observer will see a sequience of uniform outcomes either positive (UP cascade) or negative (DOWN cascade).

What would be an informational cascade in our two-sided setting?

In the terms of the strategies in one-period game ignorance of a private signal means that one uses $**AA$ or $**DD$, if the state of the world is *low* on his side, and $AA**$ or $DD**$, if it is *high* (that is one-sided cascade in BHW model). However, that might not be enough to obtain a sequience of uniform outcomes (which is the primary interest in these models).

There is a difference between the emergence of UP and DOWN cascades in our two-sided model. The difference arises due to the asymmetry in how *Agree* and *Decline* actions are translated into outcomes.

DOWN cascade Once one of the sides starts to reject the deals regardless of their signals, the outcomes will be negative. Indeed, while the opposite side might still change their beliefs, the side that is rejecting the deals has nothing to learn from the results of the negotiations: the outcomes would be negative whatever are the private signals and the actions chosen by the other side. As a result, all followers on this side will find themselves in the same situation (with the same beliefs) as their predecessors, and should reject deals regardless to their signals. Therefore, we can say that one-sided DOWN cascade ($**DD$ or $DD**$) would lead to two-sided DOWN cascade (a never ending series of negative outcomes). Since we know that outside of the regions defined by conditions in Table C.3 only equilibria with one of the sides playing $DDDD$ exist, once the beliefs leave those regions we can say that two-sided DOWN cascade emerges.

UP cascade The things are diferent for UP cascades. There are two reasons for that. First, though one of the sides may stick to *Agree* actions, the outcomes might still be negative, if the other side is not in an UP cascade. It would mean that the other side is still learning, and unlike in the case of DOWN cascades this learning may change the story.

For example, consider the case of $E = h$, $V = H$ and assume that ($AAAA$, $ADAD$) equilibrium is played. Although entrepreneurs are in an UP cascade, venture capitalists are not. Beliefs of entrepreneurs do not change, while venture capitalists' beliefs that $E = h$ strengthen when the outcome is positive or weaken when they see a failure of negotiations (see update rules C.5). Therefore, an unlucky sequence of negative signals to venture capitalists' might result in that venture capitalists will be locked in a negative cascade, hence to two-sided DOWN cascade.

Second, we have the striking possibility that even if both sides are in a positive one-sided cascade, it may still not be enough for a two-sided UP cascade. Consider an example where $E = l$, $V = L$ and ($ADAA$, $ADAA$) is played. Both sides are playing the strategies which assign them to play *Agree* regardless of their signals. Still this is not an UP cascade. Players on each of the sides hope that the state of the world on the other side is *high*. In this case the $ADAA$ decision rule implies that the other side follow its signals. Hence negative outcomes are expected from time to time, but what agents see is a long lasting series of positive outcomes. It must lead them to the suspicion that the other side is playing not AD but AA. This would imply that the true state is *low*, and consequently they might end up in a DOWN cascade instead of an UP cascade. One can also see this from the updating rules for ($ADAA$, $ADAA$) equilibrium.

We can also remark that the latter example demonstrates what we call gradual 'revelation'. One can note that ($ADAA$, $ADAA$) is a separating equilibrium, where the behaviour of the player depends on the state of the world on his side, and with

the time one can infer what is the state of the world on the other side, not from the behaviour of rivals from his own side as in BHW, but from the behaviour of the other side. Revelation does not always have to be gradual, e.g. in equilibrium ($AAAA$,$AAAD$) a negative outcome would immediately lead entrepreneurs to the conclusion that the prospective market for their technology is weak ($V = L$).

Thus for a two-sided UP cascade we have to require from the equilibrium strategy profile that both sides and all types of the players on each side choose to *Agree* regardless of their signals. The only equilibrium that results in an infinite sequence of positive outcomes is ($AAAA$, $AAAA$), which therefore we call an UP cascade in our two-sided POO model.

One remark on the our implementation of the simulations. For some p and q the region of equilibrium ($AAAA$, $AAAA$) may overlap with one of equilibrium ($ADAA$, $ADAA$). In this overlap we choose the equilibrium to play randomly. But since ($AAAA$, $AAAA$) does not change beliefs we will assume that ($ADAA$, $ADAA$) is always played, and UP cascade happens only if there is no equilibrium (excluding 'status quo' equilibria) other than ($AAAA$, $AAAA$).

4.5 Simulations

As do BHW, we can also compare information regimes with different degrees of limitation on information available to the public. Under *full information*, where E and V are known to the public, every agent chooses correct actions. In the *previous-signals-observable* (PSO) regime agents do not know the state of the opposite side, but can observe the signals their predecessor received. As a result, public information becomes more and more precise, and soon the system will converge to the optimal outcome. In the *previous-actions-observable* (POA) regime, where signals are kept private, but agents can observe actions of their predecessors, informational cascades appear, and there is always non-zero probability that the system may end up in the inferior equilibrium. Now the question to be answered is whether further restrictions on public information as POO regime will further decrease social welfare by increasing the probability of a suboptimal outcome?

We have simulated the POO model for p and q ranging from 0.5 to 1 and have estimated the probability of two-sided cascades. For each values of p and q (on a grid with step size of 0.01) we made 10000 runs. In all series the system converges to one of the cascades. The probability of a cascade was estimated as the share of realizations in our simulations that have been locked into that cascade. We use POA model as benchmark, probabilities of cascades in the POA model are given by equations (4.1) and (4.2). We found no qualititive differences in the results, neither for a higher number of runs, nor for finer grids.

Results of simulations

Our primary interest is the probability of informational cascades that lead to convergence to suboptimal choice. For this reason, we will constrain ourselves to two cases: 'negative' cascades when $E = h$ and $V = H$, which is an 'unjustified crunch', where wide-spread negative feeling about the state of the world, resulting from the chain of the unlucky events, inhibits the diffusion of the good technologies; and 'positive' cascade when $E = l$ and $V = L$, which is a socially undesirable 'two-sided bubble'.

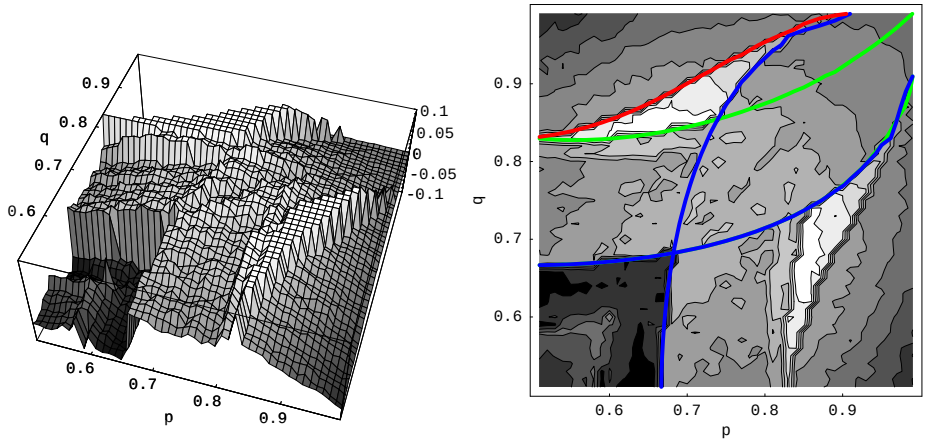

Figure 4.2: The difference between the probabilities of DOWN cascade in POO and POA models, $E = h$, $V = H$. **Left:** 3D plot. **Right:** contour plot.

Case *E=h, V=H* The difference between the probailities of 'incorrect' cascade (DOWN cascade in this case) in POO model estimated from our simulations and one of 'benchmark' POA model is presented in Figure 4.2.

Notice that for most values of p and q (white region in the density plot) the probability to be locked in the incorrect cascade is higher under POO regime than under POA regime (statistically significant). As private precision of the signals (p and q) is increasing, the difference levels out.

Another interesting feature of the Figure 4.2 is the deep 'valley' at small p and q (< 0.66). The depth of the valley is about 20%. In contrast with what has been discussed above, in the valley the probability to be locked in DOWN cascade is lower under POO regime, than under POA regime.

There are also two steps by the sides of the Figure 4.2. In the depth of those steps, as in the valley, the probability of DOWN cascades is also lower under POO than under POA regime.

Case *E=l, V=L* Figure 4.3 shows the difference in the probability of UP cascade under POO and under POA regimes. As in the previous case, the probability of

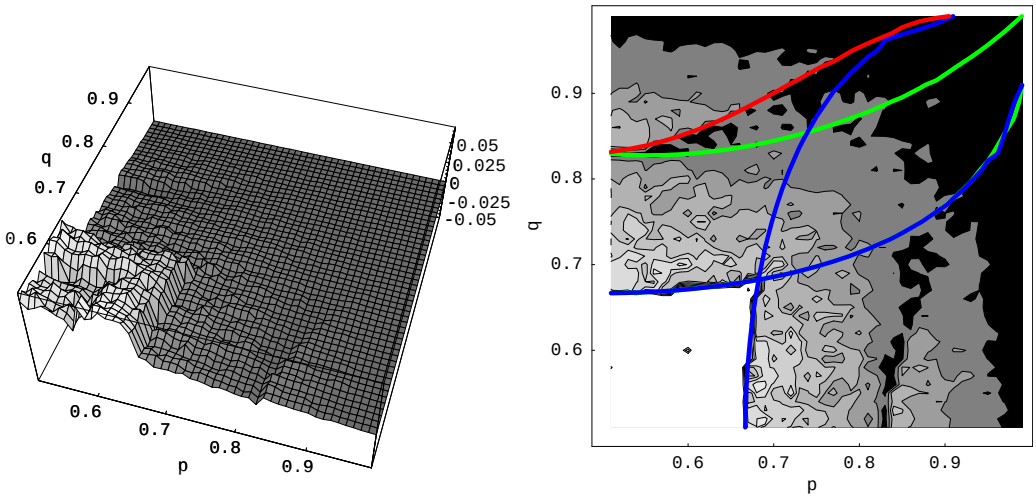

Figure 4.3: The difference between the probabilities of UP cascade in POO and POA models, $E = l$, $V = L$. Left: 3D plot. Right: contour plot.

incorrect cascade (UP cascade in this case) in POO model exceeds one of POA model (about 5% at the top of the hill at $p = 0.5$ and $q = 0.5$). Now, it does so for all p and q.

As in the case of $E = h$, $V = H$, the difference in the probabilities of UP cascades for small values of p and q is substantially higher ('island' at p, $q < 0.66$) and it is decreasing with rising p and q.

4.6 Discussion

As we have anticipated, results of our simulations for the POO model go in line with the result of BHW: restrictions on the information available to public reduce social welfare (in terms of the probability of suboptimal outcome), especially when the private information is only an inferior substitute to public.

The other conclusion from our simulations is rather striking: under the POO regime the agents seem to be 'overoptimistic': when the quality of the private signals is low the system is more prone to UP cascades.

To understand why there is 'overoptimism' in our models let us consider 'minimum series': the shortest series of outcome which result in a cascade. In the POA model similarly to BHW, a cascade starts when the two first pairs receive their positive (for UP cascade) or negative (for DOWN cascade) signals independent of

p and q. In terms of the equilibria of the one-period game it means that starting from the third pair we move from ($ADAD$,$ADAD$) to ($AAAA$,$AAAA$) or to ($DDDD$,$DDDD$).

What might be different in POO model?

Notice that when ($ADAD$,$ADAD$) is played, an agent observing a positive outcome unambigiously infers that the players must have received positive signals, i.e. a positive outcome is as informative as the private signals. It means that in the POO model as in POA, two positive outcomes should be enough to start an UP cascade.

However, observing a negative outcome one cannot unambigiously conclude which side rejected the deal, i.e. a negative outcome is less informative than corresponding actions (and private signals). Moreover, the amount of information that an agent can extract from a negative outcome depends on the accuracy of private signals on the other side as well as on the state of the world on his side.

Let $E = h$, and consider an entrepreneur who observes a negative outcome. He wants to know V, therefore he is interested which signal the entrepreneur in that pair received. If the quality of the private information on the other side, q is high, then it is less likely that the venture capitalist received a negative (wrong) signal, thus the deal has to be rejected by entrepreneur, who must have got a negative signal. If, on the other hand, q is low, so it is quite probable that venture capitalist got negative signal, then the outcome is not very informative to entrepreneurs. Thus we might expect that low precision of private information makes agents more tolerant to negative outcomes.

Where lies the border between 'high' and 'low' the precision of the signals? To answer the question we should examine the minimum series. Suppose that the first two deals have failed. Should the third pair follow their signals or join the 'herd'? Using the updating rules for $(ADAD, ADAD)$, and the equilibrium conditions, we can find p and q for which the strategy profile $(ADAD, ADAD)$ remains the equilibrium, that is, for which a cascade has not yet started, and conditions for which a cascade will certainly start ($A_2 \geq \frac{p}{1-p}$ and $C_2 \geq \frac{q}{1-q}$). The two conditions are given by:

$$\left(\frac{1-(1-p)q}{1-pq}\right)^2 \geq \frac{p}{1-p}, \qquad \left(\frac{1-p(1-q)}{1-pq}\right)^2 \geq \frac{q}{1-q} \tag{4.3}$$

(the conditions on B_t and D_t corresponding to $E = l$ and $V = L$ are not binding in this case). Figure 4.4 represents the inequalities in equation (4.3) graphically. The first inequality is represented by the light grey line labelled 1; the second by the dark grey line labelled 2. Below line 1 and left of line 2, the strategy $(ADAD, ADAD)$ remains the equilibrium even after two failures. Outside that area, after two failures a DOWN cascade surely starts.

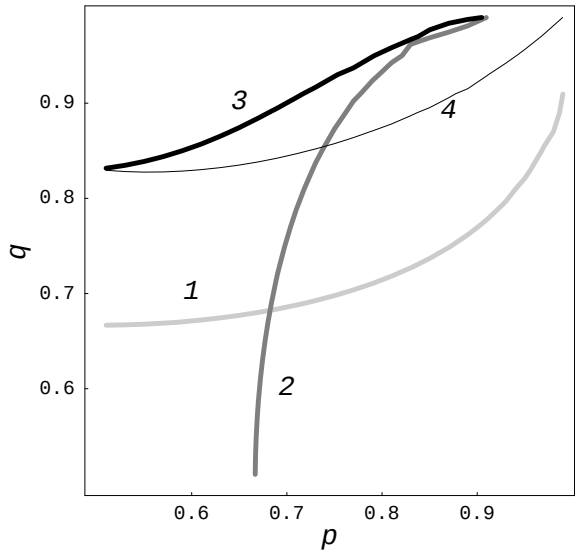

Figure 4.4: Conditions for minimum series.

Examining one other sequence helps to understand the landscapes of Figures 4.2 and 4.3. Consider that in the first four meetings there are two negative and two positive outcomes, and that currently the equilibrium strategy is $(ADAD, ADAD)$. Applying again the updating rules and the equilibrium conditions permits us to determine the conditions under which the next equilibrium strategy is $(ADAD, AAAA)$. This is the case in the region above the black line (labelled 3) in Figure 4.4, and here an UP cascade starts on one side. By contrast, for (p, q) lying between lines 1 and 4 in the figure, the next equilibrium in strategies is $(ADAD, ADAA)$ and no cascade has yet begun ($E = h$, $V = H$).

These minimum sereies results connect well with the results of the simulations. The left diagram in Figure 4.2 combines a contour plot for the case $E = h$, $V = H$ and lines of Figure 4.4. As one can see, the valley falls exactly in the range of the values for which after two negative outcomes the players are still following their signals. The two steps of Figure 4.2 are located in the area where two negative and two positive outcomes result in ($ADAA$,$AAAA$) or ($AAAA$,$ADAA$) equilibria and so on.

As one can see the main feature of Figure 4.3, the 'island' at small p and q, can also be explained by the minimum series. What is the intuition here?

One can find that after two negative outcomes neither B_2, nor D_2, which are the likelihood ratios of the agents when $E = l$ and $V = L$, are large enough to abandon conditions for 'follow-your-signal' equilibrium ($ADAD$,$ADAD$). Why, then, for p and q lying outside the island does a DOWN cascade start? As we already know, outside the island $A_2 \geq \frac{p}{1-p}$ and $C_2 \geq \frac{q}{1-q}$ and therefore if one of the sides of

the market had been in the high state it would have played $DD**$. The only reasonable reply to this strategy would be a rejection of any deal: $DDDD$. As a result, a DOWN cascade emerges. On the contrary, inside the island the agents continue playing ($ADAD$,$ADAD$) and there is still a chance to be locked into an UP cascade.

Thus, no matter what is the true state of the world in the POO model, in the situation where the quality of the private information is low, the agents tend to be 'overoptimistic' as compared with the POA regime. They put relatively more weight on positive outcomes and less weight on negative ones.

The region of low values of p and q is of particular interest in the context of the financing of the new technologies (it corresponds to top-right quadrant of Figure 4.1). In this case the 'information and trust gap' mentioned earlier is wide, because neither do investors know the technology well, nor do entrepreneurs, who are often former (or still active) academic researchers with no prior experience in business, have sufficient knowledge of the market.

Overoptimistic bias of individuals in an uncertain environment is well-known in the field of cognitive science. There are studies documenting overconfidence among entrepreneurs (Busenitz and Barney 1997) and venture capitalists (Zacharakis and Shepherd 2001). Not contesting explanations of this phenomenon from the point of view of cognitive phychology, we conjecture that two-sided interaction and information constraints discussed in this paper might contribute to overconfidence of the agents in real economy.

The recent 'dot com' crash raises questions of why the market overvalued many 'new economy' companies with immature products which had vague market perspectives. As we have seen, low quality of information (precision of private signals and incompleteness of information in public domain) results in an 'overoptimistic bias' in interpretation of the history: successful deals get more weight than failures, and agents tend to overvalue the performance of their counterparts. Overoptimism based on mutual illusions makes the system more vulnerable to two-sided 'high-tech bubbles'.

Two reservations should be made here. The first one concerns with the time scale. The model of informational cascades assumes that the ex-post payoffs are never known to the public. It is reasonable to believe that this holds true in the short-term, especially considering that the typical time between initial investment and commercialization of the new technology is several years. However in reality everlasting suboptimal herding is rather unlikely. Thus while overoptimism might have contributed to the emergence of the 'dot com' bubble, there must be other mechanisms supporting the persistence of a 'bubble'.

We shall also mention that although the overoptimism might lead to a 'wave' of investments in inferior technologies, at the same time it helps to overcome the 'information and trust gap'. If we believe that the suboptimal herding does not last

(too) long, and with the time the true state of the technology and market will be gradually revealed, the overoptimism might have a positive effect in the long-term, because otherwise the true potential of a radical innovation (probably high) might not be revealed at all. "It is better to try and fail, then never try".

Summarizing, we conclude with the following. First, dynamics of two-sided cascades in information structure where only the history of outcomes (rather than history of predecessors's actions) are observable is non-trivial and can be characterized by interactions between the two sides of market arising from learning. In comparison with POA model, where actions are public information, in POO model, although with some exceptions, the probability to end up in socially inferior cascade is higher. Second, in the situation where precision of the private signals is low for both of the sides of the market, agents tend to be 'overoptimistic' about the state of the world.

4.7 Conclusions

We examine a model in which agents on the both sides of a market have different information sets and are subject to information cascades. We assume some restrictions on available information: instead of observing actions of their predecessors as in one sided information cascade models, agents observe only successes or failures of negotiations. The changes in the information structure lead to increasing probability of locking in socially inferior informational cascade. The results support the general conclusion that can be drawn from literature about information cascades: information structure does matter, and the more restrictions on publicly available information are imposed, the higher is the probability that collective behavior will be suboptimal. Another finding of the paper is that in uncertain environment agents tend to be overoptimistic about the state of the world, which fits with results of empirical studies of financing new technologies. Overoptimism based on mutual illusions makes the system vulnerable to two-sided "high-tech" bubbles, and may be one of the reasons behind "dot com" crash.

Chapter 5

Social Emulation and Diffusion of Consumer Good Innovation.[1]

5.1 Introduction

Individual ability to imitate is not specific to human beings, but is an intrinsic feature of all social animals. However emulation as a *social process* in human societies in many respects is strikingly different emulation in animal world, as, on the one hand, no other species match humans with respect to complexity and sophistication of social organization, and on the other hand, the social structure and the institutions on which this structure is built have their mark on all social processes. The process of emulation is no exception.

In this chapter we will discuss how two particular characteristics of the society, class structure and social norms related to conspicuous consumption, shape the process of product innovation. Using a simple evolutionary model of diffusion of a positional good we show that a society with more equal class structure and social norms encouraging behavioural variety is characterized by both high market penetration and high speed of diffusion. We relate these results to the debate on the importance of demand factors in the Industrial revolution, and argue that the changes in the social structure and social norms regarding consumption prior to the industrialization in western Europe prepared the ground for introduction of the industrial methods of production.

In the next section we review relevant literature and give an overview on the framework guiding our formal analysis. We then advance the evolutionary model to study the diffusion patterns resulting from different social settings. In the fourth section we discuss assumptions and limitations of our model.

[1]This chapter is based on Reinstaller and Sanditov (2005)

5.2 Background

5.2.1 Demand as a factor in the industrial revolution

The Industrial Revolution has been studied intensively for several decades, yet some gaps in our understanding of the causes and factors of this major economic phenomenon still persist. One of the long standing issues is the importance of demand factors. According to some students of the Industrial Revolution the increasing demand for new industrial goods was an important factor in setting economies of Western Europe on the path of industrialization (Hartwell 1965, e.g.). In this respect, early work of Elisabeth Gilboy (1932) should be mentioned. She wrote

> Changing consumption standards, the increase of population and shifting of individuals from class to class, and rise in real income provided a stimulus to the expansion of industry which must not be underestimated. (Gilboy 1932, cited in Hartewell 1965)

Futhermore,

> The factory could not become typical until demand had been extended throughout the entire population to consume the products of large scale industry In order that a shift in the demand schedule may occur, individuals must be able to buy more units of commodity at the same price, or the same amount of commodity at the higher price the entire schedule must shift upwards, indicating a greater buying power.(Gilboy 1932, cited in Mokyr 1977)

Gilboy's idea that the push toward industrialization may stem from the demand echoes earlier writings of Adam Smith on the expansion of markets as a stimulus for increasing division of labour and consequent shift to industrial production

> As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or in other words, by the extent of the market. When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he has occasion for. (Adam Smith, "Wealth of Nations" Book1, ch.3)

However the "Gilboy hypothesis" received a strong critique in later works of some economic historians, of which the most influential was Joel Mokyr's (1977) "Demand vs. Supply in the Industrial Revolution". He pointed out that

> A shift of demand curve for manufacturing goods can occur only if income rises, the price of manufacturing goods falls, or if a change of tastes occur.

> Ruling out the latter for the moment, the shift of demand curve must be caused by a rise in real income, and can therefore not serve at the same time as an explanation of it.

He has examined set of theories proposed to explain the growth of demand for manufactured goods in Britain (agricultural growth, expansion of foreign demand, and population growth, demand-induced technical change, and long-term aggregate demand) and concluded that none of them provide sufficient ground to believe that the demand for industrial products might have significantly expanded during the Industrial Revolution. He has concluded that, "the traditional notion that supply and demand were somehow symmetric in the industrialization process is unfounded. The determination of "when," "where," and "how fast" are to be sought first and foremost in supply, not demand related processes"(Mokyr 1977).

Mokyr's critique effectively "suppressed" search for demand-based explanations of the Industrial Revolution for the next decade. However as more studies of the period have been made and new data on contemporary consumption have been accumulated, it has become clear that demand factor might had a significant impact on the shift toward industrial production. The new evidences suggest that despite adverse effects of stagnant or decreasing real wages, demand for manufactured goods in western Europe have been strong in 18th and 19th centuries. For example, studies of probate inventories clearly show the surge in the variety of market-supplied commodities in command of the households during (and, in fact, prior to) the industrialization phase (de Vries 1993).

The period of the Industrial Revolution in Western Europe was the time of major changes in the society, much broader than simple introduction of new production techniques. To explain the controversy between the new evidences on rise of the material culture (what de Vries called 'ever-multiplying world of goods') on the one hand, and adverse economic factors (such as low real wages) on the other hand, some economists turn their attention to the domains of the society other than those directly related to production.

Voigtländer and Voth (2005) examined importance of demographic and policy factors. They built a stochastic model of "big-push" in an economy with three sectors: agriculture, manufacture, and sector of intermediate inputs.[2] Consumers are characterized by hierarchical preferences with food produced in agriculture as a primary need, and industrial products as higher needs. In these settings unequal distribution of income consumption of industrial goods is limited by purchasing power of poor. Under favourable conditions income of poor rises that stimulate manufacturing sector and, in turn, increases demand for intermediate goods stimulating industrialization in the sector of intermediate inputs (they mention adoption

[2] A "big-push" model is immune to earlier critiques, as it does not require persistent positive demand factor, a temporary favourable demand shock might do.

of steam engine vs. traditional wind and water power). Once technical change happens in the intermediate sector the economy is set on the path of self-sustained expansion. They calibrated the model with the historical data for England and conclude that combination of relatively generous welfare system (Old Poor Law) and low pressure marriage pattern in England significantly increased the chances of industrialization in response to favourable income shocks.

Jan de Vries (1994) proposed to search the reconciliation of the controversy between Mokyr's argument and the new evidences by taking into account major changes at the level of households, the changes which, in fact, had happened *before* the industrialization started. He argued that there had been a shift of preferences from home-produced commodities toward market-supplied products. As a result, labour shifted from in-house production to the external labour markets with greater participation of women and children. It, in turn, changed the balance of power within the households making a wife more central with respect to making about spending, altering household demand in favour of clothing and consumer goods for home. In this way, even if the real wages are stagnant, the household income may grow, but even more important, the change of households organization at the aggregate level would look like a shift of preferences of a "representative consumer" in favour of market-supplied commodities (the possibility mentioned, but not examined in Mokyr's paper (1977)).

In this chapter we propose another explanation to the shift in preferences toward the market-supplied commodities. We explain the increase in the demand for manufactured goods as an effect of changes in the social norms and social structure.

David Landes (1969) discussing the causes of Industrial Revolution in Britain stated that expansion of the domestic market for industrial products had become possible due to low barriers between social classes in Britain (relative to the rest of Europe). He wrote

> Defoe's reference to the Englishman's 'expensive, generous, free way of living' calls to mind a final aspect of the British domestic market: a consumption pattern favourable to the growth of manufactures. More than any other in Europe, probably, British society was open. Not only was income more evenly distributed than across the Channel, but the barriers to mobility were lower, the definitions of status looser. (Landes 1969, p.48)

In this regard, he compared contemporary accounts for social structure in Britain and in the rest of the continent. While for Britain, the Gregory King's or Joseph Massie's schemes of the social classes described "congeries of occupational groups ranked according to wealth and so intermingled as to preclude the drawing of horizontal status lines across the whole of the social pyramid", other European countries at the time had traditional, rigid, clearly defined social classes, where one's status depended on the birth, rather than on wealth, or even less on personal abilities.

The social mobility associated with such openness of the British society was an important force in forming new consumption patterns in favour of manufactured goods.

> Mobility in such society is a force for standardization. For mobility implies emulation, and emulation promotes the diffusion of patterns of expenditure throughout population. Where there is no movement between status groups, clear, inviolate distinctions of dress and the way of life mark gradations of hierarchy. Where there begins to be movement, as in the late Middle Ages, sumptuary laws are often needed to keep people in their place. (Landes 1969, p.50)

It has to be understand, however, that, as it is often the case, a formal law is a function of underlying social norms. Hence legal regulations losing public support become obsolete and either gets transformed into exotic rituals or abolished, as it, indeed, had happened with medieval British sumptuary laws

> In England, sumptuary laws were dead letters by the end of sixteenth century; they were repealed by James I in 1604. Over the next two centuries, the trend toward homogeneity of expenditure - the effacement of vertical regional differences as well as horizontal social distinctions - continued. (Landes 1969, p.50)

According to Landes *emulation* is the force that worked in the direction of adoption of the new consumption standard. Emulation is constrained by the structure of the society and social norms. This bring us to the subject we study in this chapter: the relationship between characteristics of the society (norms and class structure) on the one side and characteristics of the diffusion process (speed, and saturation level) on the other side.

5.2.2 Social structure and social emulation

In no other sphere of economic activities the mechanism of social emulation mentioned in the previous section reveals itself so evidently as in consumption of consumer goods. The acquisition and accumulation of goods is driven not only by physical needs, but by "the recognition and admiration of our fellow human beings", as "to deserve, to acquire, and to enjoy, the respect and admiration of mankind, are the great objects of ambition and emulation", (A.Smith, cited in Rosenberg (1968, p.365)).

Commodities do not only have an intrinsic value in use, they have also a social meaning. Sociologists have long stressed that individuals value goods because they define their social position in relation to associates in lower or higher status positions. This comparison enters in the assessment of their well-being (see Bourdieu

(1984)). In a similar line Amartya Sen (1985, p.7) has argued that commodities have functionings which allow people *to do* something or *to be* something. These functionings are different from having a good or from having utility. They depend on the evaluation of the circumstances of life of a person and are also determined through interpersonal comparison.

Thorsten Veblen in his *Theory of the Leisure Class* emphasised the role of consumption as an essential tool to prove one's social status in a capitalist society (Veblen 1921). Veblen asserted that the primary function of consumption directed toward accumulation of wealth is to acquire higher status in the society:

> The motive that lies at the root of ownership is emulation; and the same motive of emulation continues active in the further development of the institution to which it has given rise and in the development of all those features of the social structure which this institution of ownership touches. The possession of wealth confers honour; it is an invidious distinction. Nothing equally cogent can be said for the consumption of goods, nor for any other conceivable incentive to acquisition, and especially not for any incentive to accumulation of wealth.

Wealth, however must be visible to the others, and therefore "to gain and to hold the esteem of men it is not sufficient merely to possess wealth or power. The wealth or power must be put in evidence, for esteem is awarded only on evidence" (Veblen 1921). As in a capitalistic society direct personal interactions are less common, this becomes obvious to others only through the consumption pattern of relevant others. Accordingly the functionings a person is able to attain through the acquisition and consumption of commodities act as important signalling devices for status. Hence commodities become the resource with which the competition of individuals for the scarce resource "status" takes place (Campbell 1995, p.104).

However, to be considered as a credible signal conspicuous consumption has to be costly to emulate by lower strata of the society:

> The consumption of luxuries, in the true sense, is a consumption directed to the comfort of the consumer himself, and is, therefore, a mark of the master. Any such consumption by others can take place only on a basis of sufferance. In communities where the popular habits of thought have been profoundly shaped by the patriarchal tradition we may accordingly look for survivals of the taboo on luxuries at least to the extent of a conventional deprecation of their use by the unfree and dependent class.

Modernization of the social norms tends to eliminate prohibitive social norms ("taboos" in Veblen's terminology). Partial emulation becomes possible, and that leads to creation of decency standards of consumption:

> Since the consumption of these more excellent goods is an evidence of wealth, it becomes honorific; and conversely, the failure to consume in due quantity and quality becomes a mark of inferiority and demerit.

The decency standards are formed by the upper class, the consumption pattern of the higher society becomes a subject of desire of the lower classes:

> The leisure class stands at the head of the social structure in point of reputability; and its manner of life and its standards of worth therefore afford the norm of reputability for the community. The observance of these standards, in some degree of approximation, becomes incumbent upon all classes lower in the scale. In modern civilized communities the lines of demarcation between social classes have grown vague and transient, and wherever this happens the norm of reputability imposed by the upper class extends its coercive influence with but slight hindrance down through the social structure to the lowest strata. The result is that the members of each stratum accept as their ideal of decency the scheme of life in vogue in the next higher stratum, and bend their energies to live up to that ideal. On pain of forfeiting their good name and their self-respect in case of failure, they must conform to the accepted code, at least in appearance.

The abolition of the norms prohibiting emulation such as medieval sumptuary laws mentioned above creates tension within the upper class, as they their consumption be distinct from the consumption of the lower classes. It provokes alteration of the consumption standards of the higher classes, but the new demarcation line will be broken again sooner or later. The dialectic tension between aspiration and distinction gives rise to a never ending race, as commodities that, at one time may confer status loose their significance once the other classes have caught up. In this way tastes do "trickle down" from the higher classes to lower social strata. This is a powerful engine of social and economic change. Nevertheless, the extent to which this will be possible will also depend on group specific factors.

The primary focus of Veblen's theory is the *inter-class* aspect of conspicuous consumption. However the *intra-class* aspects are by no means insignificant. Social norms establish behavioural regularities to which the members of a group are supposed to adhere. As consumption is a signalling device for social status, it is naturally constrained by them as well. Posner (1997) defines social norms as a rule that is based on some socially shared belief on how people ought to behave but is not promulgated by any official or legal source. They are sometimes self-enforcing, sometimes enforced by expressions of disapproval, ridicule, ostracism or codes of honour and related actions. Social norms are a behavioural public good to which every member should make a positive contribution. If that happens the behaviour is reciprocated while deviations from established patterns of behaviour are likely to be heavily punished (see e.g. Fehr and Gächter (2000, p.166)).

These mechanisms determine the pressure towards uniformity in groups. People cannot easily avoid them due to their inherited position in social space, as repeated social interactions are socially localized. Accordingly people tend to choose similar commodities than their peers because "joining the 'herd' makes their choice act less assertive and perspicuous" (Sen 1997, p. 751). Social norms determine the bandwidth within which discrepancies in behaviour are allowed and hence the ease to break with the closer social environment. Any break with the social group assigned by birth was impossible and could only happen at the danger of being marginalized by society. In more subtle ways such norms exist still today and are a defining moment of any society.

As Akerlof (1997) has stressed interactions in a social group are not only synergetic but very often they are also conflictual. People tend to move out of a group which does not share their basic values and the group in turn supports their exit in order to maintain its inner cohesion. Economic success or educational achievement may endow members of a social group with some upper-class power or attributes so that in their aspiration to a higher standard of living they break with their social environment. People who already have a high status in turn may feel the need to overcome their inherited social past, as they resent the social eminence of their peers and search for alternative means of expression. These forces give rise to behavioural variety within groups which consists of compliance with and rejection of given lifestyles.

This discussion suggests that the social factors influencing the adoption of a positional good may be condensed to effects existing between members of different social classes, namely aspiration and distinction, and intra-group effects consisting of snobbism or individualistic behaviour and conformism or bandwagon behaviour. In the model that follows we take into account these two characteristics of the social structure in which an individual is embedded to study their influence on the speed of market penetration of the new commodity. These social characteristics are the engine driving the dynamics in the model. As will be shown, different constellations of social coherence in the social classes and class homogeneity in society give rise to different patterns of diffusion.

5.2.3 Previous works

There have been few attempts to integrate social factors into (neoclassical) demand theory. Duesenberry (1949), Leibenstein (1950), Hayakawa and Venieris (1977) have tried to endogenize preferences through introduction of social or cultural propensities into the consumer's choice problem or the incorporation of Veblenian topics into utility functions. These authors imply that preferences for conspicuous or status goods are driven by comparison of individuals with social reference groups. People may react positively to the consumption pattern of some groups and adversely to

others. Accordingly, wants are shown not to be randomly distributed throughout the society but to cluster for specific social groups. Furthermore, several authors have developed status game models to study the property of demand schedules under conspicuous consumption and implications for taxation (Corneo and Jeanne 1997a, Corneo and Jeanne 1997b), as well as possible market failure resulting from it and conditions under which it can be avoided (Pesendorfer 1995, Bagwell and Bernheim 1996). The diffusion of new products is not an explicit aim of the analysis of these papers.

Other work has partly addressed this question. Some authors have shown that if the behavioural patterns of an individual are alternatingly enforced or dampened by the behaviour of significant others, chaotic demand patterns may emerge (see, for instance, Congleton (1989), Iannaccone (1989) or Rauscher (1993)). Similarly Cowan, Cowan, and Swann (1997) have formulated a stochastic model whose dynamic is based on aspiration-, bandwagon- and Veblen effects. They show that if certain consumer groups seek distinction and others aspire to their behaviour cyclical consumption patterns and consumption waves may emerge. In a similar fashion Janssen and Jager (2001) explain market dynamics with lock-in, fashions or unstable renewal. They posit that it is dominated by the behavioural rules of consumers reflecting preference for distinction or conformity.

Cowan et al. and Janssen and Jager identify consumption norms as an emergent property of systems of single agents interacting with others, out of personal preferences for snobbism, aspiration or distinction. While we believe that social norms do indeed emerge from social interaction of individuals, i.e. that they are endogenous, it is also the case that in the short run individuals act embedded in an already given social structure. Social norms, for instance, lead often to institutions which have a semi-permanent character and change only slowly as norms change. In a short run analysis therefore the social structure should be assumed to be given as it constrains and determines the choices of agents. For this reason, instead of taking an agent based view, we pursue a population approach and study adoption patterns and behavioural variety as well as the speed of diffusion in dependence of a given social structure.

5.3 The model

We formalize the considerations put forward in the previous section as an evolutionary game with two groups of individuals.[3] As such our model is concerned with the frequency evolution of consumption strategies in the economic system. The members of each population are heterogeneous. They are assumed to be boundedly

[3]See Taylor (Taylor 1979), as well as Cressman (Cressman 1995) and Weibull (Weibull 1995) and the classic references cited there.

rational and to use simple consumption routines, a positional good, to signal social status. Each of the routines is a pure strategy in a game. The pay-offs derived from consumption are based on the individual's social position and the characteristics of the population as a whole. The agents in the game get pay-offs from consumption out of the interaction with the members of their own social class, as well as from the interaction with members of the other social classes. Based on these pay-offs the agents in each population learn which routines improve their well-being. The learning process is captured by replicator equations.

The model is analytically solvable. Its equilibria represent stable norms of consumption to which the economy converges after an innovation has been introduced. We establish the local stability of these equilibria, but unlike most theoretical work in evolutionary games the focus of our model does not lie on the investigation of the application of the evolutionary stable strategy solution concept to the dynamic stability of the replicator dynamic. We focus on the study of the adjustment process towards an equilibrium once a new commodity has been introduced into the economy. For this purpose we simulate the behavior of the model for some limit scenarios and analyze the resulting diffusion patterns.

5.3.1 Players and Strategies

There are two social classes in our model. Class structure is captured by an "average" available consumption budget per unit of time and individual $w_i, w_1 < w_2$ in each class i (i=1,2), and the share of a social class in the total population $q_i, q_1 > q_2$ with $q_1 + q_2 = 1$. As the model is set up, the share of the lower class, q_1, is a measure of the distribution of social attributes: if $q_1 = 0.5$ we have a perfectly equal class structure, as the two cohorts are of equal size. One should be careful that this does not imply that income distribution is even as well. Indeed, for the parameter settings used throughout this paper it implies an unequal income distribution. The social attributes delimiting the class boundaries may be thought of as cultural and social capital. The former depends on the family background as well as investments in and commitments to education that underlie academic attainment, while the latter represents class specific stocks of social trust and norms as well as networks that people are embedded in and use. Both together define the identity and the boundaries of a social class. In the model population shares and available income (w_1, w_2, q_1, q_2) are exogenously given and are assumed to be constant over time.

We assume that at each unit of time an individual of a population chooses a consumption basket consisting of two parts: a positional or status good (we will use the terms interchangeably), and a basic good. The pay-off from consuming a basket is the sum of the pay-offs from consuming its parts. In choosing a status good an individual has the choice between two alternatives: good X with price p_x, or good Y with price p_y. We assume that the endowment of each individual is large enough to

consume any of the two commodities, $(w_i > p_y, p_x)$; $i = 1, 2$. The positional goods are indivisible. They may be thought of as a bundle of complementary commodities that are all needed to pursue a certain lifestyle. For simplicity we also assume that they have no other value than a social one. In other words, individuals derive value from owning the good or not as this conveys social status and not from its intrinsic use value. Consumption of the basic good in turn has no particular social meaning and has decreasing marginal value in use. The basic good is perfectly divisible. We assume that prices reflect marginal costs, and that they do not change over time.[4]

The share of the lower class consuming good X at time t is given by x_1. The remaining share of the population consuming good Y is $y_1 = 1 - x_1$. Similarly, x_2 and $y_2 = 1 - x_2$ are the shares of the upper class of individuals consuming X and Y, respectively. In terms of the model these goods are pure strategies or routines played by the agents to signal status. The strategies are denoted as e_x when she chooses the basket with good X and e_y when he goes for the basket with good Y. The population states for the two classes are then defined by $\mathbf{s_1}$=(x_1, y_1) for the lower class, and by $\mathbf{s}_2$=(x_2, y_2) for the upper class.

5.3.2 Payoffs

Positional or status goods Capturing some of the considerations advanced in the previous section, we assume that consumption of goods X and Y is driven purely by positional or status considerations. We analyze two important behavioral motives described before: *aspiration* and *distinction*. The lower class aspires to the standards of decency demonstrated by the upper class, the social elite. Members of this group would like to signal status similar to that of the upper class. In terms of our model this means that they want to buy what the upper class buys. On the other hand the upper class seeks distinction to preserve their status as social elite.

As for the behavior within the two social groups we will examine a spectrum of different "social norms" ranging from a society forcing strict behavioral compliance on their members, to a "non-conformist" society, where it is important for any individual to emphasize his own identity and individuality from the others. In a conformist society mechanisms of retaliation will sanction deviant behavior. Conversely in a "non-conformist" society individuals "aping" others will be perceived as a nuisance and accordingly retaliatory mechanisms will ensure that this does not happen too frequently.

We assume that an individual is engaged into two contests per unit of time against a randomly drawn opponent from the total population. The choice of a specific consumption profile depends on his expected payoff in this matching. We

[4]This is done for analytical clarity. The overall results do not change if we assume falling prices (due to scale or learning economies) for the new good.

Table 5.1: Payoff matrix: **D** (distinction).

	e_x	e_y
e_x, (basket with good X)	-1	1
e_y, (basket with good Y)	1	-1

choose the following specification of the payoff matrix to formalize the "distinction"-effect (Table 5.3.2).

An agent gets a positive pay-off whenever his choice is different from his opponent's choice, and nuisance (of the same magnitude) when the choices coincide. In the same way, matrix **C** enables us to capture conformist behavior. Here, conversely as in the case of "distinction" behavior, an agent gets positive payoff if he chooses the same basket as his opponent and experiences a negative outcome otherwise (Table 5.3.2).

Table 5.2: Payoff matrix: **C** (conformity.)

	e_x	e_y
e_x, (basket with good X)	1	-1
e_y, (basket with good Y)	-1	1

Basic good The consumption of basic goods does not produce any particular social signal. We include them to capture the reproductive aspects of the social system as a whole, i.e productive consumption, which transcends class boundaries and is not related to social demonstration effects. The consumption of basic goods depends on imperatives in daily life associated with a wide variety of factors other than social signalling. They reflect the needs of individuals in a specific historical context and therefore they have social meaning, but do not produce social signals. They are complementary to class related lifestyles, which are driven by the social relations enshrined in the class structure.

We assume that every part of income that is not spent on the positional good is spent on these "basics". The marginal use value of consuming $w - p$ of the basic good is decreasing, i.e. the value of the income not spent on the status good falls on the margin for higher incomes. To capture this standard assumption we use - without loss of generality - a concave function given by $\sqrt{w - p}$.

A basket An individual consuming a basket that contains basic and positional goods gets a payoff which is given by the sum of the pay-off from the consumption

of the basic good and the expected payoff from consuming a status good with social meaning. For an agent out of the low income group playing strategy e_k, $k = x, y$ the expected payoff is given by

$$u_1(e_k; \mathbf{s}_1, \mathbf{s}_2) = q_1 \Big[\mathbf{e}_k \left(\omega \mathbf{C} + (1-\omega) \mathbf{D} \right) \mathbf{s}_1 \Big] + q_2 \Big[\mathbf{e}_k \cdot \mathbf{C} \mathbf{s}_2 \Big] + \mathbf{e}_k \cdot \mathbf{w}_1, \tag{5.1}$$

where $\mathbf{e}_k$ is a unit vector in $\mathbb{R}^2$, and $\mathbf{w}_i$ is the vector of pay-offs of the basic goods

$$\mathbf{w}_i = \begin{pmatrix} \sqrt{w_i - p_x} \\ \sqrt{w_i - p_y} \end{pmatrix},$$

with $i = 1, 2$. The first term in equation (5.1) describes interactions within the lower class. Here the parameter $\omega \in [0, 1]$ captures the "social norms" in place in a society. We assume that $\omega = 0$ for a perfectly "non-conformist" set-up, while $\omega = 1$ in the case of strictly conservative social norms. All the values within these limits represent the more realistic intermediate cases expressing a tendency to conformity if $\omega > 1/2$ or to individualism if $\omega < 1/2$. A special case worth mentioning is where $\omega = 1/2$. At this point the intergroup heterogeneity of each population vanishes and the game transforms into a contest between two homogeneous populations. The second term in equation (5.1) arises from the *aspiration* effects present in the social class and finally the third term is just the use value derived from consumption of the basic good.

From equation (5.1) we derive the average payoff for the lower class, which is given by

$$u_1(\mathbf{s}_1; \mathbf{s}_1, \mathbf{s}_2) = q_1 \Big[\mathbf{s}_1 \cdot \left(\omega \mathbf{C} + (1-\omega) \mathbf{D} \right) \mathbf{s}_1 \Big] + q_2 \Big[\mathbf{s}_1 \cdot \mathbf{C} \mathbf{s}_2 \Big] + \mathbf{s}_1 \cdot \mathbf{w}_1. \tag{5.2}$$

The pay-offs from consumption to individuals of the upper class is derived in a similar fashion. In analogy to equation (5.1) equation (5.3) defines the pay off for an individual playing strategy e_k which is

$$u_2(e_k; \mathbf{s}_1, \mathbf{s}_2) = q_1 \Big[\mathbf{e}_k \cdot \mathbf{D} \mathbf{s}_1 \Big] + q_2 \Big[\mathbf{e}_k \cdot \left(\omega \mathbf{C} + (1-\omega) \mathbf{D} \right) \mathbf{s}_2 \Big] + \mathbf{e}_k \cdot \mathbf{w}_2. \tag{5.3}$$

Given equation (5.3) the average payoff in the upper class is

$$u_2(\mathbf{s}_2; \mathbf{s}_1, \mathbf{s}_2) = q_1 \Big[\mathbf{s}_2 \cdot \mathbf{D} \mathbf{s}_1 \Big] + q_2 \Big[\mathbf{s}_2 \cdot \left(\omega \mathbf{C} + (1-\omega) \mathbf{D} \right) \mathbf{s}_2 \Big] + \mathbf{s}_2 \cdot \mathbf{w}_2. \tag{5.4}$$

The second and third terms of equation (5.3) are similar to the ones in equation (5.1) for the lower class. The difference is the first term which captures the wish of the upper class to distinguish themselves from members of the lower class.

5.3.3 Replicator Dynamics

The replicator dynamics captures the learning behaviour in each population. The market share of each good increases or falls depending on whether agents discard or adopt a good that allows them to seek social status based on their past experiences. We use the standard two-population replicator dynamics introduced by Taylor (Taylor 1979) to analyze our model. The dynamics is defined by the system of four differential equations

$$\begin{aligned} \dot{x}_i &= x_i \cdot \Big[u_i(e_x; \mathbf{s}_1, \mathbf{s}_2) - u_i(\mathbf{s}_i; \mathbf{s}_1, \mathbf{s}_2)\Big] \\ \dot{y}_i &= y_i \cdot \Big[u_i(e_y; \mathbf{s}_1, \mathbf{s}_2) - u_i(\mathbf{s}_i; \mathbf{s}_1, \mathbf{s}_2)\Big] \end{aligned} \tag{5.5}$$

with initial conditions $x_i(0) = 1-\varepsilon,\ y_i(0) = \varepsilon$, i=1,2. In the appendix to this paper we examine the local stability of the replicator dynamics for the equilibria towards which the model converges under different parameter settings. The model is not stable in a small domain of the parameter space, as shown in Figure 5.1. This is discussed in detail later.

5.3.4 Equilibria of the model

The equilibria in this model represent stable norms of consumption towards which the system gravitates once an innovation has entered the economy. They indicate how much behavioural variety is present in an economy in dependence of a given social structure.

There are four types of possible equilibria in the model: two in pure strategies (pooling and separating), one in mixed strategies and one where the upper class plays pure (consuming the new good), while the lower class plays mixed strategies. In what follows this is denoted as partially mixed strategy. Furthermore we determine under which parameter values for class structure ($q_1 > 0.5$) and social norms (ω) a given equilibrium exists and is locally stable in the replicator dynamics (5.5). This is shown in the appendix. For all parameters other than q_i and ω we use the same values as we employ for our simulations, i.e. $w_1 = 1, w_2 = 2, p_x = 0.5, p_y = 1$.

Pooling (no penetration) equilibrium: $y_1 = 0$**,** $y_2 = 0$**.**

This is an equilibrium where there is no diffusion of the new good at all. The equilibrium requires

$$\Delta u_1(0,0) < 0, \qquad \text{and} \qquad \Delta u_2(0,0) < 0.$$

where $\Delta u_i = u_i(e_y; \mathbf{s}_1, \mathbf{s}_2) - u_i(e_x; \mathbf{s}_1, \mathbf{s}_2)$, $i = 1, 2$ represent gains or losses in pay-offs for each of the classes from consuming the new or the old good. According to (5.1)

and (5.3) $\Delta u_i(y_1, y_2)$ is then defined as

$$\begin{aligned} \Delta u_1(y_1, y_2) &= 4\alpha q_1 y_1 + 4q_2 y_2 - \Delta w_1 - 2\alpha q_1 - 2q_2, \\ \Delta u_2(y_1, y_2) &= -4q_1 y_1 + 4\alpha q_2 y_2 - \Delta w_2 + 2q_1 - 2\alpha q_2, \end{aligned} \tag{5.6}$$

where $\alpha \equiv 2\omega - 1$ and $\Delta w_i \equiv \sqrt{w_i - p_x} - \sqrt{w_i - p_y} > 0$. Substituting expressions for Δu_i from (5.6) we get

$$q_1 < \frac{2 + \Delta w_1}{2(1-\alpha)}, \qquad \text{and} \qquad q_1 < \frac{2\alpha + \Delta w_2}{2(1+\alpha)}. \tag{5.7}$$

When these inequalities hold, there is no diffusion. Accordingly, the share of the new good in equilibrium will be zero,

$$Y^* = 0. \tag{5.8}$$

Separating equilibrium: $y_1 = 0$**,** $y_2 = 1$**.**

Under social set-ups leading to separating equilibria only the upper class adopts the new positional good, while the lower class uses the old one. The conditions for the equilibrium are

$$\Delta u_1(0, 1) < 0, \qquad \text{and} \qquad \Delta u_2(0, 1) > 0.$$

or by substituting as before

$$q_1 > \frac{2 - \Delta w_1}{2(1+\alpha)}, \qquad \text{and} \qquad q_1 < \frac{\Delta w_2 - 2\alpha}{2(1-\alpha)}. \tag{5.9}$$

The equilibrium market share of the new good when inequalities (5.9) hold is given by

$$Y^* \equiv q_1 y_1 + q_2 y_2 = q_2. \tag{5.10}$$

Partially mixed equilibrium: $0 \leq y_1 \leq 1$**,** $y_2 = 1$**.**

The third possible equilibrium is a partially mixed equilibrium where the upper class uses only the new positional good, while the lower class uses both. The conditions for the equilibrium are

$$\Delta u_1(y_1^*, 1) = 0, \qquad \text{and} \qquad \Delta u_2(y_1^*, 1) > 0. \tag{5.11}$$

Accordingly, from the first equation in (5.11) we can find the equilibrium market share for the lower class y_1^*:

$$y_1^* = \frac{1}{2} + \frac{\Delta w_1 - 2q_2}{4\alpha q_1} \qquad \text{for} \quad \alpha \neq 0. \tag{5.12}$$

The equilibrium market share for the upper class, if condition (5.11) is to hold, is $y_2^* = q_2$. From this together with (5.12) the market share of the new good in equilibrium is given by

$$Y^* = q_1 y_1^* + q_2 = 1 - \frac{q_1}{2} + \frac{\Delta w_1 - 2q_2}{4\alpha} \qquad \text{for} \quad \alpha \neq 0 \tag{5.13}$$

To determine the domain of q_1 and ω, where the equilibrium exists we substitute y_1^* into inequality (5.11) and in addition we require $0 < y_1^* < 1$. It gives us

$$\begin{aligned} q_1 &< (>) \quad 1 - \frac{\Delta w_1 + \alpha \Delta w_2}{2(1+\alpha^2)}, \\ q_1 &< (>) \quad \frac{2 - \Delta w_1}{2(1-\alpha)}, \\ q_1 &> (<) \quad \frac{2 - \Delta w_1}{2(1+\alpha)}, \end{aligned} \tag{5.14}$$

for $\alpha > 0$ ($\alpha < 0$).

Equilibrium in mixed strategies: $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$.

Finally, there is an equilibrium in the model, where individuals from both classes use the old and the new positional good. It must hold that $u_{1x} = u_{1y}$ and $u_{1x} = u_{1y}$. This implies that the following conditions hold,

$$\Delta u_1(y_1^*, y_2^*) = 0, \qquad \text{and} \qquad \Delta u_2(y_1^*, y_2^*) = 0.$$

The solution to the system of equations is

$$\begin{aligned} y_1^* &= \frac{1}{2} + \frac{\alpha \Delta w_1 - \Delta w_2}{4(1+\alpha^2) q_1}, \\ y_2^* &= \frac{1}{2} + \frac{\Delta w_1 + \alpha \Delta w_2}{4(1+\alpha^2) q_2}. \end{aligned}$$

Therefore the market share of the new good is

$$Y^* = q_1 y_1^* + q_2 y_2^* = \frac{1}{2} + \frac{(1+\alpha)\Delta w_1 - (1-\alpha)\Delta w_2}{4(1+\alpha^2)} \tag{5.15}$$

The solutions of the system must be in the range (0,1). From this follows that equilibrium (5.15) exists in the area in parameter space delimited by the following conditions

$$\begin{aligned} q_1 &> \frac{\alpha \Delta w_1 - \Delta w_2}{2(1+\alpha^2)}, \qquad & q_1 &< 1 + \frac{\Delta w_1 + \alpha \Delta w_2}{2(1+\alpha^2)}, \\ q_1 &> -\frac{\alpha \Delta w_1 - \Delta w_2}{2(1+\alpha^2)}, \qquad & q_1 &< 1 - \frac{\Delta w_1 + \alpha \Delta w_2}{2(1+\alpha^2)}. \end{aligned} \tag{5.16}$$

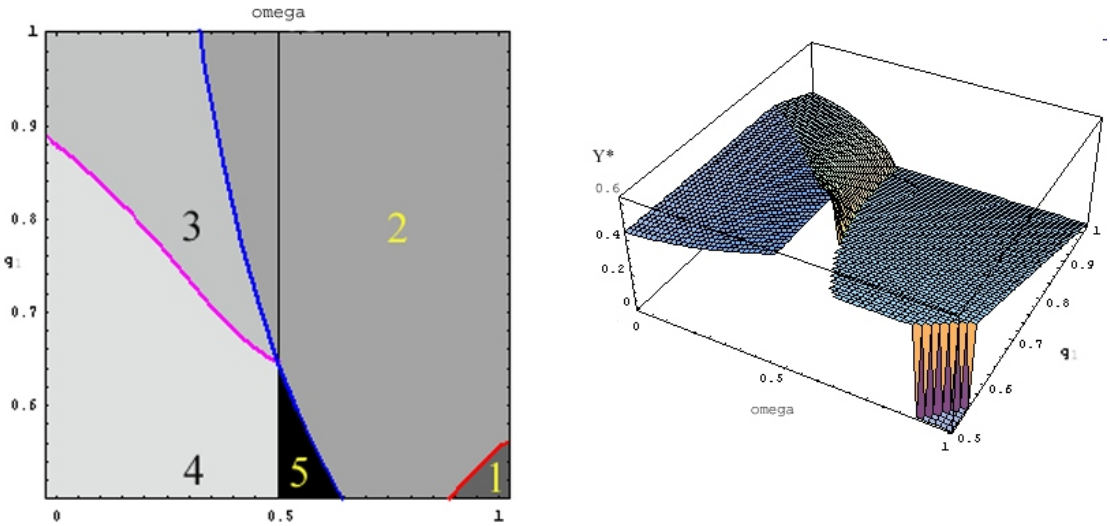

Figure 5.1: **Left:** Domains of the stable equilibria of the model. (1) Pooling (no-penetration) equilibrium ($y_1 = 0$, $y_2 = 0$), (2) separating equilibrium ($y_1 = 0$, $y_2 = 1$), (3) partially mixed strategy equilibrium ($0 \leq y_1 \leq 1$, $y_2 = 1$), (4) mixed strategy equilibrium ($0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$), (5) unstable equilibrium. **Right:** Equilibrium market shares over the parameter space.

Domains of the equilibrium and stability

The domain of the parameters for social norms and class structure (ω, q_1) $(0 \leq \omega \leq 1$, $0.5 \leq q_1 \leq 1)$ for which the different equilibria exist and are locally stable can be divided into five parts by inequalities (5.7), (5.9), (5.14), (5.16) and the stability conditions given in the appendix. This is depicted in the left part of Figure 5.1.

For combinations (ω, q_1) enclosed by area 4 only mixed equilibria exist and are stable. As the distribution of social characteristics gets more unbalanced, we move to area 3 where the new good Y is consumed by all individuals from the upper class and by some individuals of the lower class. Here we observe partially mixed equilibria. If the social norms change towards conformism, we step into the domain with separating equilibria 2, which are stable all over area 2. The no penetration (pooling) equilibrium domain 1 is located in the bottom-right corner of Figure 5.1. In 1 both no penetration and separating equilibrium exist and are stable, therefore the replicator dynamics given in equation (5.5) may converge to any of them depending on the initial conditions. For the initial conditions used in our baseline model we observe convergence to the pooling equilibrium. A mixed strategy equilibrium exists for the parameter values enclosed by area 5 however it is not stable and shows limit cycles oscillating between on the boundary (i.e. $Y^* = 0$ and $Y^* = 1$). The reason for this is that at $\omega = 0.5$ the intra-group effect of the game vanishes and becomes a game between homogeneous populations. A glance at matrices $\mathbf{C}$ and

D shows that matching strategies are of opposite sign. Thus, independent of the consumption strategy a member of the distinction group chooses *a-priori* she has always an incentive to switch to the alternative strategy if she is matched with a member of the conformist group playing the same strategy. For these solutions an analysis of the diffusion patterns is not meaningful.

The right part of Figure 5.1 shows the equilibrium market shares for the new luxury good introduced into the economy over the equilibrium domain shown in the left part of the figure. For parameter constellations covered by area 1 the market share of the new positional good is zero as there is a pooling equilibrium. In what corresponds to area 2 the equilibrium market share is given by separating equilibria, it therefore falls with the share of the upper class given by q_2. In area 3 (partially mixed equilibria) we observe higher market shares for the new status good as compared to the situation where the social norms lead to a separating equilibrium in area 2. The market share of the new luxury is nevertheless highest in the domain where both classes use both goods, given by area 4. Finally, in area 5 there is no stable equilibrium, as the share of the new goods fluctuates between values close to zero and one. This area is left blank in Figure 5.1.

5.3.5 Analysis of the diffusion paths for some limit cases

We use our model for the analysis of the role of class structure and social norms in the process of diffusion of a new good. We start with simulations over some limit cases capturing perfect conformity $\omega = 1$ and total non-conformity $\omega = 0$, as well as set-ups for a society with equally distributed social characteristics and a society with an uneven distribution of social characteristics, i.e. for parameter values $q_1 = 0.5$ and $q_1 = 0.9$ respectively. We assume that the consumption prior to a date $t = 0$ is limited to only one basket with good X. At that moment in time a new product Y is introduced with an initial market share ε (for both classes). The price of the new product, p_y, is higher than the price of the old one, p_x. We examine the model for $p_x = 0.5, p_y = 1$ and consumption budgets $w_1 = 1, w_2 = 2,$. The initial market share of Y for both social groups is set to $\varepsilon = 0.01$. This initial condition captures a situation where a previous luxury has turned into a common good, so that members of the upper class are not able to distinguish themselves from lower class people.

We use the Fisher-Pry substitution rate (Fisher and Pry 1971) to captures the speed at which an old commodity is driven out of the market. This model suggests a logistic substitution trajectory

$$\frac{y}{1-y} = \exp(a + bt).$$

In its linear transform $ln(\frac{y}{1-y})$ the slope b of a fitted straight line captures the learning or substitution speed, while the intercept a measures the adoption delay.

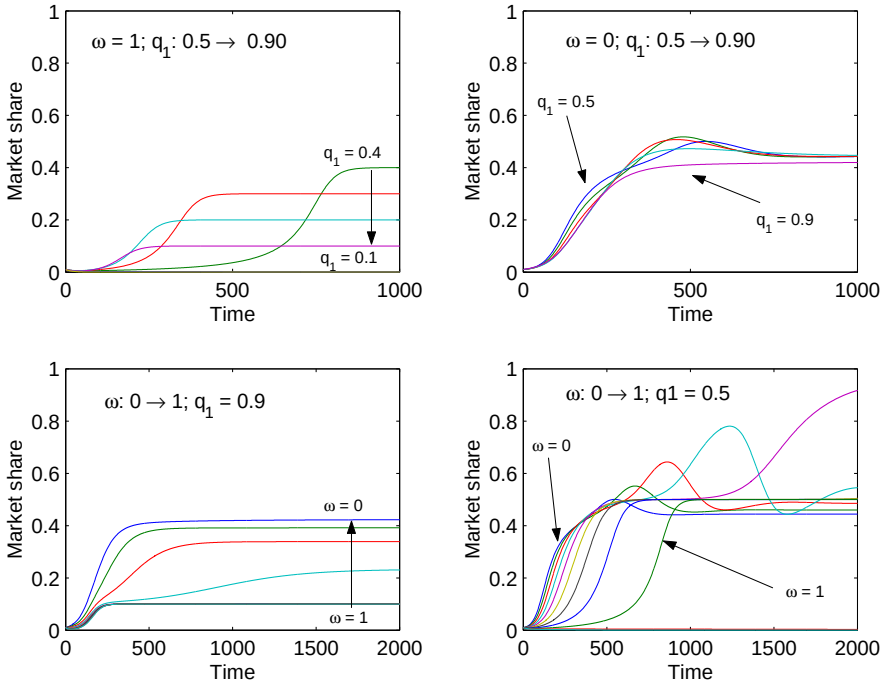

Figure 5.2: Diffusion paths, examples of limit cases. Top left: perfect conformity, effect of equality. Top right: non-conformity, effect of equality. Bottom left: unequal class structure, effect of non-conformist social norms. Bottom right: even class structure, effect of non-conformist social norms (some solutions with $\omega \approx 0.5$ give rise to oscillations, i.e. "fashion cycles").

The steeper this line is, the faster substitution takes place, the larger the intercept (in absolute terms), the higher is the adoption delay or consumer inertia. The intercepts and slopes of simulated diffusion data fitted to the Fisher-Pry equation are presented in Figure 5.3.

Effect of an unequal distribution of social attributes

We first examine two polar cases of social norms and analyze how changing class structure influences the diffusion pattern of the new consumer good introduced at time $t = 0$. The paths our model generates for these constellations are shown in the upper quadrants of Figure 5.2.

Conservative social norms, $\omega = 1$ Under perfect conformity the intra-class distinction effect is not present, i.e. the $\mathbf{D}$ matrices in the terms capturing inner-group interaction in our equations (5.1) and (5.3) disappear. In this case people of

the same social class who are randomly matched play a coordination game given by pay-off matrix **C**. The top left quadrant of Figure 5.2 shows the diffusion curves resulting for a parameter range $0.5 \leq q_1 \leq 0.95$. The effect of growing equality in the distribution of social characteristics is twofold: on one hand it speeds up diffusion and on the other hand the level of saturation falls.

With the given initial conditions at $q_1 = q_2$ the members of the lower class derive the same expected pay-off from being equal to their peers as well as from being equal to the upper class. There is no incentive for members of the lower class to switch to the new commodity, as given the initial market shares of the new commodity in the two groups, the pay-offs are already almost at their highest level. We observe at first a pooling equilibrium given by equation (5.8). The few initial adopters will switch back to the old positional good due to existing peer pressure.

A change in equality in the distribution of social characteristics has the effect on the upper class to increase negative pay-offs from being equal than the lower class. It starts paying upper class individuals to adopt the new status good. The model settles on separating equilibria given by equation (5.10). Due to the pressure towards conformity adoption is very slow. Diffusion takes longest under near-equality conditions (see Figure 5.2) but eventually the whole upper class will adopt the new positional commodity.

Non-conformist society, $\omega = 0$ The top right quadrant of Figure 5.2 indicates that the diffusion paths resulting in a non-conformist society are quite different. In a perfectly non-conformist social constellation the matrix **C** capturing intra-group interaction in equations (5.1) and (5.3) vanishes. In this case people of the same social class who are randomly matched play a "hawk-dove" game given by pay-off matrix **D**. Each individual seeks to be different from its peers. This means that over the parameter range of q_1 up to the value of $q_1 = 0.9$ the model settles on a mixed strategy equilibrium given by equation (5.15) for both social classes, and to a partially mixed equilibrium as in equation (5.13) at that value and beyond. This implies that under the "individualistic" setting of this run the equilibrium reflects an economy with maximum variety on the market for most class structure parameters. With fixed ω the market share for the new good gravitates around 0.45, and falls in the partially mixed equilibrium range (see Figure 5.1, right).

The inter-group effects are responsible for short fashion waves visible as an overshooting over the final saturation level. As the frequency of adopters of the new good in the lower class increases, the upper class starts perceiving negative pay-offs from buying it, while the pay-offs for the lower class increases, so that there is an incentive to adopt more of it. This triggers some members of the upper class to revert to the old commodity. With the frequency of adopters of the new good in the upper class decreasing, the pay-offs from consuming the new good fall for indi-

viduals of the lower class as well and the model settles on the mixed equilibrium. When the opportunity parameter is changed towards values capturing inequality in the distribution of social characteristics, the upper class will restrict consumption of Y earlier as lower class members are encountered at higher frequency, thus causing the overshooting to appear earlier. Rising inequality in the distribution of social characteristics has the effect to dampen out fashion waves as the parameter range for partially mixed equilibria is approached.

Effect of conformity

While in the first set of runs we examined the diffusion path along the vertical parameter axis of the equilibrium domain in Figure 5.1, we now change parameters to move along its horizontal parameter axis, examining how changes in social norms influence the process of diffusion for any given class structure. The diffusion curves are shown in the quadrants at the bottom of Figure 5.2. The diffusion paths resulting from these model runs are hybrids of the first two cases studied so far.

Unequal distribution of social characteristics, $q_1 = 0.9$ In this case we observe partially mixed Nash equilibria and separating equilibria, as the parameter for conformity is changed from 0 to 1. At low conformism players in both populations would tend to use both goods but as the inequality in social characteristics is high and the probability for upper class types to encounter similar lower class types is high, it pays them to play a pure strategy, even though it may cause negative pay-offs when playing against peers. The shift of ω towards conformity leads to a fall in the adoption of the new good in the lower class as the pay-offs to individuals being equal to their peers starts outweighing snobbism and aspiration effects. The saturation level shifts downwards and the speed of diffusion decreases. The upper class on the other hand continues to have an incentive to adopt the new commodity due to the high frequency of members of the lower class in the total population. The market share drops to the share of the upper class.

Equal distribution of social characteristics, $q_1 = 0.5$ In changing the parameter ω over its parameter range the model settles on four possible equilibria. Under non-conformity we observe mixed strategy equilibria, in the parameter range of $0.5 \leq \omega < 0.6$ the model exhibits a cyclical behavior, beyond that separating equilibria and close to perfect conformism there is a pooling equilibrium. Whether the model settles on the latter depends on the initial conditions chosen and this is the case for the parameter value for ϵ we use.

Fashion cycles emerge as the intra-group effects vanish. Members of the lower class start deriving higher pay-offs from being equal to members of the upper class,

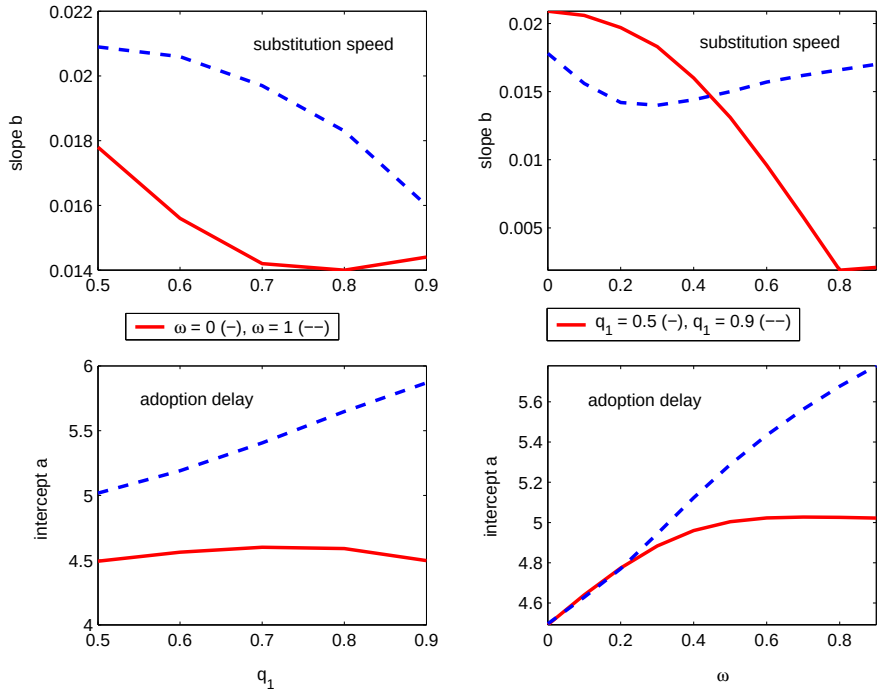

Figure 5.3: Parameters of fitted Fisher-Pry substitution curves. The parameter a captures the adoption delay, while the parameter b captures the learning or substitution speed.

while the latter's negative pay-off increases through this development. In approaching the critical value dampened cycles appear, which converge to a stable saturation level after some time. At $\omega = 0.5$ nevertheless, intra-group effects completely vanish and fast cycles emerge. The upper class has a continued incentive to change its consumption pattern, as the lower class catches up. Only after conformity becomes stronger it pays better for members of the lower social class to stick on the same consumption pattern as their peers and forgo pay-offs from imitating the upper classes.

Market penetration and diffusion speed

The right part of Figure 5.3 shows the intercept values and slope of the linear substitution curves fitted to the data generated from the runs with changing distribution of social characteristics. They reflect the adoption delay or inertia and the substitution speed. The first is clearly higher for conformism than for non-conformism and tends to increase with an increasingly unequal distribution of social characteristics, while the latter is faster under conformism and is falling with increasingly unequal

distribution of social characteristics. Over the parameter range of ω instead the adoption delay is practically equal for parameter values capturing non-conformism but while it levels out in the equality scenario it increases steadily in the inequality scenario. This picture is reversed for the substitution speed. In the part of the parameter space where the adoption delays are equal for the equality and inequality scenarios it is faster for the equality scenario but falling as parameters are set to capture conformism.

Table 5.3: Summary of the results

	Equal	Unequal
Conformist	$\prec$	
	$\asymp$	$\asymp$
Non-conformist	$\succ$	

These results are summarized in table 5.3. The market penetration and diffusion time is slowest under a conformist setting with equal distribution of social characteristics. A social set up with conformism and unequal distribution of social characteristics fares better. These two set-ups in turn perform worse than non-conformist ones in terms of diffusion speed, but also in terms of saturation market shares given in the right part of Figure 5.1. Hence we find that the social set-up with non-conformism and even class structure leads to a faster diffusion of new consumer goods than all other social constellations.

5.4 Discussion

Conspicuous Consumption and Technology: Diffusion of Clock One of the central assumptions of the model is that diffusion of new technology may be driven by conspicuous consumption. Is such an assumption grounded?

An interesting evidences in support of this hypothesis can be found in the fascinating history of diffusion of clock described by Carlo Cipolla in his book about "Clocks and Culture". The clock and horology made a significant contribution to the rise of modern industrial economy. Lewis Mumford ("Technics and Civilization", 1934) stated 'the clock, not the steam engine, is the key machine of the modern industrial world'.

Although the first European public clock had been invented already in Middle Ages, diffusion of clock was very slow, as clock was very expensive in those days. It was costly to build and expensive to maintain. The decision to install or not to install a clock was often the result of long and heated debates. The utilitarian aspect of

having a public clock in the town was certainly significant but not always definitive argument in those debates. In those cases it often happened that "conspicuous public consumption" was a decisive factor for ordering a clock:

> Some towns rivaled others for the distinction of having, as a fifteenth-century French document put it, '*relotgium magnum sufficiens et honorabile ad honorem villae*'. About 1380, the Town Council of Lyon decided to install on one of the bridges a tower with a clock similar to one on a bridge in Paris: '*prout et quemadmodum edificate sunt Parisiis turris et horlogium desuper existens*'. In the 1420's the Town Council of Romans (France) decided to build a very beautiful clock 'without any regard to expenses' (*'sans regarder 'a la depense'*). In 1557 the inhibitants of Monte'limar (France) decided to have a clock similar to that of romans: '*a la forme d'icelluy de Romans*'.

In this way conspicuous consumption led the diffusion of the clock despite its high cost.

The conspicuous nature of the public clock had an interesting effect on the direction of the development of clockmaking technology. As it is easier to impress public by appearance of the clock than with precision of its work, towns and cities rivalry expressed itself in ordering clocks with rich ornamentation, or sophisticated mechanically animated performances rather than in paying clock masters for inventing more precise time keeping mechanisms

> The most striking occurrence in the early history of clocks is that while medieval craftsmen did not improve noticeably in precision, they soon succeeded in constructing clocks with curious and very complicated movements ... For the sake of beauty and civic pride, complicated movements were sometimes added to existing clocks.

Evolution of personal clocks and watches has followed a similar pattern: lot of clockmakers effort was spent on the appearance rather than on precision.

The story of clock also sheds some light on some other aspects of the relationship between demand and supply. Production of domestic clocks started almost as early as production of public clocks.[5]. However in that period demand for clock was limited by nobility (and only by its most educated part). Even in the largest cities of Europe (such as Paris) the demand was not large enough for craftsmen to specialize in horology. Thus in the fourteenth and fifteenth centuries clocks were made by blacksmiths, gunfounders, and locksmiths. It started to change as the demand for clocks diffused down into the social pyramid. In the sixteenth and seventeenth

[5] "When that extraordinary collector of *objets d'art* Charles V of france died in 1380, the officials who made the inventory of the 3,985 items of his collection found among them 'one clock all made of silver and with no iron, that belonged to the late King Philip the Fair, with two weights covered with silver and filled with lead'. " (Cipolla 1967)

centuries domestic clocks and watches became less of a rarity, and "[b]eing luxury items they were at the very centre of the craze for exuberant decoration that characterized the late Renaissance and the Baroque Age." At this point growing demand permitted the formation of settled groups of clockmakers (in perfect accordance with Adam Smith's argument about the role of demand in division of labour mentioned earlier).

Furthermore, the history of clockmaking illustrates the relationship between science, technology, and the role of demand. As clock become more popular it has attracted attention of scientists:

> In the Renaissance, while clocks and watches were becoming progressively more fashionable in the upper class as useful and graceful ornaments, the clock as a machine attracted more and more the inquisitive curiosity of scholars, amateur scholars, and learned people in general. ... Among those devoted themselves to the problems of measuring time and of constructing accurate clocks one can mention Galileo, Christian Hygens, Robert Hooke, Godefroy Wendelin, Nicolas Fatio, and Wilhelm Leibnitz.

In this respect, increasing demand for clocks worked as a 'focusing device' (Rosenberg 1969).[6]

Thus the history of clock illustrates that, indeed, conspicuous consumption may work as a force in shaping demand for new technology. It also shows that conspicuous consumption may have effects on the development on the technology, fostering specialization, and influencing the direction of innovative search.

Limitations and direction for future research The model presented in this chapter turns out to be a useful device to link the history of technology and the history of consumption. Nevertheless, the results presented in table 5.3 on which this reinterpretation was based are less general than they may appear at first. One should be careful as consumption is a much more complex phenomenon than our model is actually able to capture.

The picture may change if social norms are no longer symmetric across classes. The rich being also the powerful may be able to evade social pressure more easily than lower class people. In this sense social norms may not be independent of class structure. One could think of situations where class homogeneity is low and variety is still high. When upper class lifestyles are too expensive so that lower class people cannot buy the goods or when the upper class is so small, that lower class people have actually no chance to get an idea how rich people live, rich eccentrics may feel compelled to be distinct not from lower classes, which is in any way the case, but from their peers. So, on the one hand the model neglects class internal distinction

[6] The next chapter has more on 'focusing devices' and technological paradigms.

processes (see e.g. Swann (2001)), while on the other hand, it also does not take into account that not all products are introduced in the upper end of the market.

There are limitations to the results concerning the speed of diffusion of new goods as well. Consumption of specific goods may generate very strong positive network effects that were not present in similar goods before, so that independently of any social consideration people may have a preference for the new. Furthermore, we assume that both new and old goods are intrinsically the same. However, the social norms for some reasons may favour one type of good over the other. In such case the pressure to conform might speed up the diffusion of the new good in contrast with our result. Finally, in many circumstances no social signal may be attached to a commodity either because its consumption is not visible to other people or because new goods make life just easier given the specific needs in a specific historical period of time. A good may be so widespread in society that nobody could infer any social signal from it. Under these circumstances diffusion would be driven by other mechanisms than those discussed here.

Finally, a model studying consumption as a pure demand driven phenomenon also neglects the effects of marketing strategies of firms, improvements in existing product and other supply driven effects on consumer behaviour. The interaction between these two domains should be studied more thoroughly in future research efforts. Even more so, as historical studies suggest that besides the continuous search for new combinations on the side of products and technology, entrepreneurship is also predicated on the creation, exploration and exploitation of new markets. Our model may turn out to be a useful starting point for that agenda as it indicates under which social conditions opportunity for firms to engage into the creative exploration of new markets may be high. Further modelling efforts should explore the interaction between product innovation and opportunity conditions generated by specific social set-ups more in depth.

5.5 Conclusions

In this chapter with a simple model of conspicuous consumption we have studied the influence of parameters reflecting social structure on the diffusion paths of product innovations of consumer goods. We used the set-up of an evolutionary multi-population model with two populations. The first population is the upper class, whose members act as innovating force in consumption. The second population is the lower class, whose members imitate the consumption behaviour of the higher class. We assumed that in both classes there are social norms exerting pressure on their members not to innovate or imitate, i.e. to develop an "individualistic" consumption behaviour. We explored the influence of changes in class structure between the two classes, as well as the effect of social norms on the speed of diffu-

sion of new products their take-off time and the market saturation level. The main result of the model is that novelty diffuses most rapidly in a social setting where class structure is equal and behavioural variety is high. This social set-up leads to a faster diffusion of novelty than in all other constellations. In other words, societies allowing for more behavioural variety and with an even class structure should experience a more dynamic consumer behaviour than otherwise. Based on these conclusions the model offers a theoretical reinterpretation of the historical record on the rise of consumerism and its relation to the literature on the role of the demand in the Industrial Revolution.

Chapter 6

Patent citations, the value of innovations, and path-dependancy

6.1 Introduction

It is well recognized these days that only efficient production, accumulation, and utilization of technological knowledge can ensure long term economic growth. Planning and implementing R&D programmes have become a routine task for many governments and companies around the world. Therefore the knowledge about the distribution of returns from R&D is of great practical importance.[1].

The main problem hindering research in this direction has been scarcity of data on R&D. However with the arrival of new data, particularly patent data, and with advances in methodology the field is rapidly expanding. The evidence accumulated in recent years confirms earlier findings and are univocal on the overall features of the distribution of the innovation values: it is highly skewed with most of the innovations having value close to zero, and few innovations scoring very high, a fact that has direct implications for planning and evaluation of innovation policies and firm strategies (Scherer and Harhoff 2000).

Although the extreme skewness of the distribution is now a well established fact, the precise form of the distribution of the innovation values is still under debate. In particular, there is a controversy about the right tail of the distribution. Based on the results of a survey of holders of German patents Harhoff, Scherer, and Vopel (1997) report that the best fit for the tail (defined as innovations with values over DM 23,000) is obtained with lognormal distribution (vs. Pareto and Singh-Maddala distributions). On the contrary, applying techniques of extreme-value theory to the set of different data on the innovation values, Silverberg and Verspagen (2004)

[1]Certain features of the distribution such as whether it has a "heavy" tail are also relevant for more theoretical research. For example, see (Jones 2005, Kortum 1997, Houthakker 1955)

demonstrate that if the lower bound of the tail is set correctly, the tail is fit better with Pareto distribution (rather than with lognormal distribution).

So far, research in this direction has been focussed on the properties of the distribution. In this paper I propose to approach the problem from the other end: instead of questioning what is the exact form of the distribution of innovation values I inquire about the *process* that generates the distribution. I will argue that the evolutionary theory of technical change is helpful in understanding the dynamics of innovation values.

My argument proceeds along the following lines. First, according to evolutionary theory the process of technical change is incremental and path-dependent, and development of a technology follows "technological trajectories". Success of an innovation, related to the resolution of an important design problem, plays the role of "focusing device" (Rosenberg 1969): it directs innovative search to the areas of "technology space" opened by the innovation, and stimulates the flow of the inventions based on the technology it represents. Following this logic we make an assumption that the value of the innovation depends on the range of the problems it can be applied to: the more general it is, the higher is its value. Combining this assumption with path-dependency in the process of technical change we expect the dynamics of innovation values to be path-dependent: the more valuable the innovation is, the more likely it is to be employed in consequent innovations, as a result, the more valuable it will become.

To formalize this intuition I propose a simple model based on generalized Polya processes that takes into account the path-dependent nature of technical change, and show that the model fits the distribution of patent citations (a measure of the value of patented inventions) very well.

The paper is structured as follows. In the next section I review the literature on how the value of innovations is related to the characteristics of the process of technical change, and the measures of the value of innovations with particular attention to patent citations. Then I formulate the model and describe the data that I use later to test the model. The results of fitting the distribution of patent citations are presented in Section 6.5. Section 6.6 discusses several aspects of the model. The last section concludes the paper.

6.2 Literature

6.2.1 Value of Innovations and Direction of Technical Change

It is almost obvious and self-evident that the value of an innovation depends on the characteristics of the technical change, and therefore the value cannot be defined out of the context of the history of the technology. This fact is not very important for a

retrospective judgement (because the history has been already realized), however if we aim at a dynamic view of technical change, then before questioning what is the distribution of the value of innovation, we must try to answer what is the *process* that makes some innovations more important than others.

The answer to this question depends on what kind of picture of the technical change one has in mind. For a simple linear model of technological progress which seems to dominate modern growth literature, the answer is straightforward: technological progress is nothing but expansion of the production set, therefore the innovation offering the highest reduction of production costs (for a process innovation), or/and higher quality of the product (for a product innovation) will have the highest value which is steadily decreasing as new, better methods of production keep coming (e.g. Aghion and Howitt 1992). Assuming that the gain in the productivity is distributed according to some probability law, the uncertainty surrounding the value of the innovation is contained in the demand and the rate at which innovations arrive. Moreover most models assume that demand does not change over time (embedded in the utility function of a representative consumer). Therefore *ex-post* distribution of the values can be inferred from the distribution of the productivity gains and the rate of the technical change. There is no place for the direction of technical progress, because in the linear model there is only one (toward increasing productivity).

On the other hand, according to the evolutionary tradition in the economics of technology, the process of technical change follows path-dependent "technological trajectories" punctuated by discontinuities of "natural trajectories"/ "technological guideposts"/"technological paradigms" (Nelson and Winter 1977, Sahal 1981, Dosi 1982). Most of the time we expect to observe relatively stable clusters of (interlinked) technologies, with more valuable core technologies in centre of each cluster.

The clustering of the patented inventions in technological space has been analyzed with the use of patent documents through IPC classification, patent citations (in bibliometric style), and textual analysis of the patent documents. Pier, Rost, Teichert, and von Wartburg (2003) use (EPO) patent citation data to decompose the "technological blob" of mobile telecommunication. Huang, Chen, Yip, Ng, Guo, Chen, and Roco (2003) use longitudinal patent data for nanoscale science and engineering to make country, institution and technology field comparisons. They employ both content map analysis and patent citations. On time-series content maps they observe several dominant topics occupying different periods of time. Graff (2003) surveys the use of patent data for identification of micropatterns in innovations in agricultural technology.

There are (at least) two factors behind path-dependency in technical change which tend to 'bunch' technologies together: (a) complementarities between contemporary technologies, and (b) localization of the search in the technological space,

due to the bounded rationality of agents.[2]

In the study of interdependencies between technologies in the American economy Nathan Rosenberg notes

> Inventions hardly ever function in isolation. Time and time again in the history of American technology, it has happened that the productivity of a given invention has turned on the question of the availability of complementary technologies. Often technologies did not initially exist, so that the benefits of potentially flowing from invention A had to await the achievements of inventions B, C, and D. These relationships of complementarity therefore make it exceedingly difficult to predict the flow of benefits from any single invention and commonly lead to postponement in the flow of such expected benefits. (Rosenberg 1982, p.56)

Rosenberg supports this thesis with a number of examples from the history of transport sector, agriculture, electricity, machine tools, metallurgy *etc.*

Silverberg and Verspagen (2005) point out that even such seemingly simple invention as a bicycle is, in fact, a collection of many related inventions including "pneumatic tyres, ball bearing (and thus precision machining, the precision grinding machine ...) [...] without which bicycle boom of the 1890s would have been unthinkable". They formulate and examine a model in which a new technology becomes feasible only if it has links with the technologies already in use. In this model the importance (value) of the technology depends not only on the increase in the productivity this technology offers, but also on whether this technology can make other technologies available, i.e. on the direction of technical change.

Another perhaps less 'visible' factor behind path-dependency in technical change is related to cognitive aspects of the innovation process. From the very beginning, modern evolutionary economics has recognized that economic agents are characterized by bounded rationality (Nelson and Winter 1982), i.e. their behaviour is governed not by full optimization over the complete set of control variables, but by the process of trial-and-error, some search heuristics (embedded in the routines of an organization), or optimization over a subset of control variables in a limited domain. A range of different models of the search process can be found in the current evolutionary economics literature (Frenken 2004, for a survey): simple trial-and-error similar to the learning in an evolutionary game, genetic algorithms when strategies are coded in binary strings, and a new string arises through recombination of parent strings (Birchenhall 1995, Dawid 1999), NK-models of search on technological landscapes, an application of the cutting-edge theories from the evolutionary biology (Kauffman 1993, Frenken and Nuvolari 2004), and simulated annealing, a method

[2]For different technologies the relative importance of systemic and cognitive factors mentioned here may differ.

of combinatorial optimization originated from modelling thermodynamic systems in physics (Cooper 2000).

It is worth emphasising that given our research question we shall view the search process not at the level of individual agents performing their search on their own, but as a process that involves the whole technological community; this community includes inventors, firms, government labs, academicians and the like. There is an obvious parallel with the sociology of science, in particular, with Thomas Kuhn's "scientific paradigms". During the stable phase of the development of a technology researchers and engineers have a number of standard approaches to solve standard problems shared by the community. To solve a particular engineering problem means finding an appropriate standard solution (design) and adjusting it to the problem (Cooper 2000).

Furthermore, the *research agenda* (i.e. what needs to be improved, what can be achieved with available techniques *etc.*) is also shared at the community level. As a result, the direction of innovative search is framed by the current state of the technology and hence depends on the previous success (in terms of both technological achievements and commercial benefits). Such a picture of innovative search goes along with the views of Rosenberg (1969, 1974) who sees inventive activities as focused on a set of related engineering problems ("focusing devices/technological imperatives") which result in "compulsive sequences" of innovations over time.

Given the path-dependent nature of the process of technical change it is reasonable to expect that the dynamics of innovation values is also path-dependent, and, indeed, in the next section I show that under some plausible assumptions this dynamics can be described by path-dependent stochastic process. We assume that the value of a given innovation depends on the generality of the design problems arising along the current "technological trajectory" to which the innovation may be applied. An innovation representing a successful solution to a set of important design problems works as a "focusing device": it attracts more innovative search in the related area of the technology space. Search aims both to improve the solution and to explore the area of the technology space opened by the innovation. Increasing intensity of the search, in turn, leads to an increasing flow of innovations based on the given innovation, and expands the scope of the problems to which the technological knowledge underlying the innovation can be applied. Therefore, other things equal, the path-dependent nature of technical change implies a path-dependent dynamics of innovation values.

Summarizing at this point, we can state that that the value of an innovation depends on the direction of technical change: if the innovation fits with the current "technological paradigm" it is likely to be used in consequent innovations, and its value will grow. Furthermore, success of particular innovations shape the direction of technical change: successful innovation becomes a "focusing device" and will be replicated time and time again either because it is a key to link new technologies

with existing ones as in the model of Silverberg and Verspagen (2005), or because boundedly rational agents use it as a starting point in the process of search in technology space.

Our assumption about the value of innovation and the direction of technical change does not contradict the intuition one can get from patent citation literature. Trajtenberg (1990) explains that "if citations keep coming, it must be that the innovation originating in the cited patent had indeed proven to be valuable". Somewhat similarly Harhoff, Narin, Scherer, and Vopel (1999) word it as " it is reasonable to suppose that the prior inventions cited in new patents tend to be the relatively important precursors that best define the state of the art. The broader the shoulders, the more likely they are to be cited". In the next section I will explain how data on patent citations can be used to quantify my argument. First, however let us review how value of patented innovations reveals itself in the patent data.

6.2.2 Patent citations and the value of innovations

There are several ways to assess the values of patented inventions. Pakes and Schankerman (1984), Pakes (1986), Schankerman and Pakes (1986) have employed data on patent renewal to estimate the characteristics of the values of the patent rights. Lanjouw, Pakes, and Putnam (1998) extended this framework in order to utilize data on the applications for a patent (related to the same invention) in different countries ("family size"). Another approach to assessing the value of patents is to use the stock market valuation of a company to which the patents have been granted (Griliches 1981, Pakes 1985, Hall, Jaffe, and Trajtenberg 2005a, among others). Yet another stream of research that has proved to be very productive is to utilize information contained in the patent documents themselves (number of citations, number of claims, number of IPC classes). In particular, citations-based indices have been very successful (Trajtenberg 1990, Trajtenberg, Henderson, and Jaffe 1997, Harhoff, Narin, Scherer, and Vopel 1999, Jaffe and Trajtenberg 2002). Finally, in the recent paper Harhoff, Scherer, and Vopel (2003) found that outcomes of opposition against patent grants proved to be highly informative for predicting the value of the patent rights (taken from a survey of holders of German patents).

Results of most studies indicate that the measures of the patent values mentioned above are mutually coherent, and more importantly, most of the measures correlate well with the value of the patents inferred from the direct surveys of the patent holders. Harhoff, Narin, Scherer, and Vopel (1999), Harhoff, Scherer, and Vopel (2003) tested a set of different measures of patent quality as predictors of the patent value obtained from the survey of holders of German patents and found that forward citations, family size, outcomes of opposition proceedings, and whether patents were renewed to a full-term correlate well with the patent values.

Among other measures of the patent quality measures based on citations (in

particular, forward citations) are appealing for a number of reasons. First, as has already been mentioned, most studies suggest that the number of citations a patent has received (*forward citations*) is a good proxy for social and private returns to the innovations. Second, all information needed for the construction of appropriate measures is contained within the patent document. Third, modern software and publicly available computer-readable data make it easy to construct these measures tailored to different patent classes, institutions, countries *etc.*[3]

Furthermore, one might interpret patent citations to prior art as "paper trails" of knowledge spillovers (Jaffe and Trajtenberg 2002). Such interpretation of patent citations led to a prolific research avenue in different areas of innovation studies ranging from spatial economics (Jaffe, Trajtenberg, and Henderson 1993), to university-industry links (Henderson, Jaffe, and Trajtenberg 1998), and social network analysis (Balconi, Breschi, and Lissoni 2004).

Nevertheless, patent citation data should be used with some caution. First, there is a problem with the "benchmarking" of citation data (Hall, Jaffe, and Trajtenberg 2001). For example, if we are to compare two patents taken in different years, and suppose that the older patent has received more citations than the other one, then it is not clear if it is because the old patent is more valuable, or simply because, it had more chances to be cited (truncation problem). It is also important to keep in mind that the stock of patents is rapidly growing; hence, other things equal, the earlier patent have higher chances to be cited, then the patents which were taken later. Furthermore, changes in practices in Patent Offices and in patenting strategies of firms may lead to additional complications for an intertemporal comparison.

Second, results of surveys of innovators have cast some doubts that patent citations represent "paper trails" of *direct spillovers* i.e. the fact that the owner of a citing patent learned about the innovation contained in the cited patent from the cited patent itself or from the holder of this patent prior to the invention (Jaffe, Trajtenberg, and Fogarty 2000). It is not rare that innovators have learned about the predecessors of their patents only at the stage of patenting. Many citations have been added by patent examiners, or innovators' attorneys, and hence cannot be regarded as an evidence of direct spillovers.

The problem with benchmarking can be resolved if we limit comparison of patents to one cohort, i.e. to patented inventions made in more or less the same time, provided that by the time of observation the patents have accumulated enough citations. This, in turn, raises a question about dating the patents. A patent document published by a Patent Office of interest, in our case - the United States Patent and Trademark Office (USPTO) for the NBER dataset, and the European Patent Office (EPO) for the CESPRI dataset, contains several dates: *priority date*

[3]It is worth to mention that the literature is virtually silent about how and why patent citations arise. A rare exception is Bertran (2003).

- the date when the patent was applied to the Patent Office in any jurisdiction; *application date* - the date when the inventor filed the documents for a patent to the Patent Office of interest; *grant date* - the date when the Patent Office issued the patent to the inventor. We are interested in the date closest to the time of invention, which is the priority date for the EPO patent data, and application date for the USPTO data[4]. Therefore to avoid the problem with benchmarking we shall select patents with the priority/application date within a small period of time (one year seems to be appropriate time span).

The concern about whether a patent citation represents a direct spillover, or it is evidence of an indirect spillover coming through the "word-of-mouth" via the social network of inventors (Breschi and Lissoni 2004), is not essential for our purposes. Indeed, the inventor of the citing patent does not have to be in a contact with the inventor of the cited patent, neither does he need to know the details of the patent, he might even be unaware that the patent exist. What he knows is that a solution to certain design problem is feasible and it exists, so that he can focus on a complementary technology. The assumption we make about the value of innovations is that the value positively depends on the number of the design problems it may be successfully applied to. As far as patent citations correctly trace the lineage of the technologies, i.e. it links related technologies and establishes the precedence of the inventions, the number of citations received by a patent is a measure of the generality of the design problem a solution to which the patent represents, and as a result the number of citations received by the patent is positively correlated with the value. Whether a citation is added by inventor, her attorney, or patent examiner is not important.

6.3 The Model

The model is based on the evolutionary view of technical change and patent citation literature outlined in the previous section. According to the evolutionary theory the value of an innovation depends on how well the innovation is embedded in the current "technological paradigm", i.e. on the frequency with which the technological knowledge underlying the innovation is utilized in the consequent development of the technology. We also assume that a citation received by a patent documents an instance when the piece of knowledge represented by the patent has been used. Thus, in accordance with the patent citation literature we can state

Assumption 1 *The value of a patented invention is reflected by the number of citations received by the patent: the higher is the number of citations, the more*

[4]NBER dataset provides no priority dates, however if we consider a cohort of patents issued to the US inventors the difference in the dates is likely to be small.

valuable the invention is.

Furthermore, a successful innovation might work as a "focusing device" for the consequent innovative search. It is reasonable to assume that the impact of innovation on the direction of development of technology depends on the current value of the innovation. The more valuable the innovation is, the more likely the particular piece of technological knowledge represented by the patent will be utilized in consequent innovations, and, as a result, the more valuable it will become. According to Assumption 1 the growing importance of the innovation will be reflected in the frequency of citations the patent will receive. Therefore,

Assumption 2 *The higher is the value of a patented invention, the more likely it is to be used by consequent innovations, the more valuable it will become, and the more citations it will receive.*

The model can be formalized as follows. Consider N patents at time $t = 0$ indexed by i, $i \in \{1, \ldots, N\}$. At time $t = 0$ the patent i has value $v_{i,0}$, reflected by the number of citations it has received, $c_{i,0}$. Each moment in time one citation is made.[5] The probability that the patent i is cited is proportional to the value of the technology the patent i represents, $v_{i,t}$

$$p_{i,t} = \frac{v_{i,t}}{\sum_{j=1}^{N} v_{j,t}} \tag{6.1}$$

A citation received by patent i implies that technology i has been used, and reflects the increase in the value of the patent, i.e. $v_{i,t+1} > v_{i,t}$ if $c_{i,t+1} = c_{i,t}+1$. We consider two "value functions" mapping values into citations: a linear function

$$v(c_{i,t}) = v_0 + c_{i,t}, \tag{6.2}$$

and non-linear function in the form

$$v(c_{i,t}) = v_0 + {c_{i,t}}^{\alpha}. \tag{6.3}$$

Inserting (6.2) into (6.1) we can rewrite it as

$$\Pr(c_{i,t+1} = c_{i,t} + 1 | c_{1,t}, \ldots, c_{N,t}) = \frac{v_0 + c_{i,t}}{V_0 + t}, \tag{6.4}$$

where $V_0 = \sum_{j=1}^{N}(v_0 + c_{j,0})$, i.e. the sum of the patent values at time $t = 0$.

For the non-linear value function (6.3) we have

$$\Pr(c_{i,t+1} = c_{i,t} + 1 | c_{1,t}, \ldots, c_{N,t}) = \frac{v_0 + {c_{i,t}}^{\alpha}}{v_0 N + \sum_{j=1}^{N} {c_{j,t}}^{\alpha}}. \tag{6.5}$$

[5] Time in this model is measured in citations. It is not the same as the calendar time.

Formulas (6.4) and (6.5) define stochastic processes that belong to the class of generalized Polya processes (finite case). Early applications of Polya processes in economics go back to the works of an IIASA group in the 1980s (Arthur, Ermoliev, and Kaniovski 1983). The recent revival of the interest in the Polya processes was induced by the rapidly growing literature on the evolution of networks originating from the studies of WWW, but spread into a number of disciplines (physics, ecology, molecular biology, sociology *etc.*).

The generalized Polya process (Chung, Handjani, and Jungreis 2003) can be defined as follows

Definition 1 *For fixed parameters, $\alpha \in R$, $0 \leq p < 1$ and a positive integer $N > 1$, begin with N bins, each containing one ball and then introduce balls one at a time. For each new ball, with probability p, create a new bin and place the ball in that bin; with probability $1-p$, place the ball in an existing bin, such that the probability that the ball is placed in a bin is proportional to c^{α}, where c is the number of balls in that bin.*

For a *finite Polya process* $p = 0$, i.e. no new bins are created. If $p > 0$ we have an *infinite Polya process*. Parameter α describes the type of feedback: it is said that there is *positive feedback*, if $\alpha > 1$, *negative feedback* if $\alpha < 1$, and *linear feedback* if if $\alpha = 1$. The case of $\alpha = 1$ and $p = 1/2$ is often referred in the literature as the *preferential attachment scheme* (Albert and Barabasi 2002, Barabasi 2002).

The infinite process with different types of the feedback function has been studied extensively in the context of network growth (mostly to explain the distribution of nodal degrees). In particular, the preferential attachment scheme has received a lot of attention. The nodal degrees of the resulting graph, so called scale-free network, are distributed according to a power (Pareto) law, which is often seen as an indication of self-organization and can be observed in nature in a variety of situations (Barabasi 2002). However for our purposes we shall limit our attention to the finite case.

For the finite process with linear feedback ($p = 0$, $\alpha = 1$) such as one defined by (6.4) it is possible to show that as time (the number of balls) goes to infinity, the proportions of the balls in the bins (*a.s.*, almost surely) approach their limits X_i, $i \in \{1, ..., N\}$, which are distributed uniformly on the simplex $\{(X_1, ..., X_N) : X_i > 0, X_1 + ... + X_N = 1\}$ (Chung, Handjani, and Jungreis 2003, Theorem 2.1).[6] It follows that, the distribution of the proportions has an exponential tail in drastic contrast with the infinite case mentioned above.

The limit distribution of proportions is different for the other types of the feedback. For negative feedback, $\alpha < 1$, balls are distributed equally among bins, i.e.

[6]The processes defined by equations (6.4) and (6.5) are not exactly the same as the process in Definition 1, because, in general, v_0 is not equal to 1. However it does not affect the results for the limit distributions mentioned here.

$X_i = 1/N$ for any $i \in \{1, ..., N\}$. If positive feedback is the case as in the process defined by (6.5), then $X_i = 1$ for one bin and $X_i = 0$ for the other balls, i.e. a "winner takes all" situation (Chung, Handjani, and Jungreis 2003, Theorem 2.2). The latter case is interesting, we may expect to see long and probably fat tails at any finite time.

The results for the *limit* distribution ($t \to \infty$) mentioned above are indicative for what we can expect for the *asymptotic* distribution, $(x_1^t, ..., x_N^t)$ for $t >> N$: in the case of linear feedback the distribution the tail of the distribution is decreasing exponentially, in case of the positive feedback we may expect to see heavier tails. However, a distribution arising at finite time (the number of citations in our case) which we are interested in can be quite different from the limit distribution.

Most studies of Polya processes focus either on the asymptotic distribution ($t >> 1$, when the initial conditions are not important) for the infinite case (Albert and Barabasi 2002, Krapivsky, Redner, and Leyvraz 2000), or on the limit distribution for the finite case and the rate of convergence toward the limit distribution (Bassanini and Dosi 1999). To the best of my knowledge there are no general results concerning the distribution at any given period of time.

In the Appendix to this paper using the rate equation describing the evolution of the distribution of the number of the balls in a bin I derive the recursive formula for the distribution at any given t. In case of the linear feedback the solution can be found in a closed form (formula (E.4)). For the case of the non-linear feedback there is no solution in the closed form and therefore we have to rely on the results of simulations.

Simulations have been carried out as follows. First, we fix a cohort of patents $i = 1...N$ and termination date, find the distribution of patent citations in the beginning of observations and at the terminal date. We also choose some values of the parameters v_0 and α. Formula (6.5) defines the probabilities that at time t patent $i = 1..N$ will be cited. According to this probabilities we randomly pick up one of the patents, let it be patent j, and update $c_{i,t}$ for the next round: we increment the number of citations received by patent j: $c_{j,t+1} = c_{j,t} + 1$, while the number of citations made to other patents remains the same: $c_{i,t+1} = c_{i,t}$ for $i \neq j$. The procedure is repeated for $t + 1$, and so on. The process starts from the initial distribution of patent ciations, and stops when the sum of the citations to the cohort reach the number of the citations at the termination date. Resulting distribution was compared with the observed distribution of patent citations.

6.4 Data

The NBER dataset created by Hall, Jaffe, and Trajtenberg (2001) contains data for all utility patents granted by the USPTO from 1963 to 1999 (about 3 million

patents) and all citations made by patents granted from 1975 to 1999 (about 16 million citations).

For my study I have chosen patents applied in 1989, similar to the cohort used in (Silverberg and Verspagen 2004). Since the NBER dataset contains no priority dates, there might be a problem with dating the patented inventions related to the patents applied earlier in other (then the USA) countries, because for these inventions the date of invention (which we are interested in) is likely to distant from the date of application to the USPTO. Moreover, patent citations might have a 'home country bias' i.e. other things equal there may be a bias towards citing the patents granted to US inventors. These problems can be reduced if we restrict our focus to the patents issued to the US inventors (first inventor), assuming that before applying to other Patent Offices US inventors are more likely to apply for a patent at USPTO. In addition, it also helps to avoid a potential complication due to home country bias. This leaves 50,263 patents in the 1989 cohort and 341,365 citations received by these patents from 1990 to (including)1999. The distribution of the citations from 1989-1999 is shown at the left diagram of Figure 6.1 (blue triangles). The distribution is highly skewed with the large share of patents having received near zero citations. It also has a long and heavy tail. The most cited patent has received 245 citations.

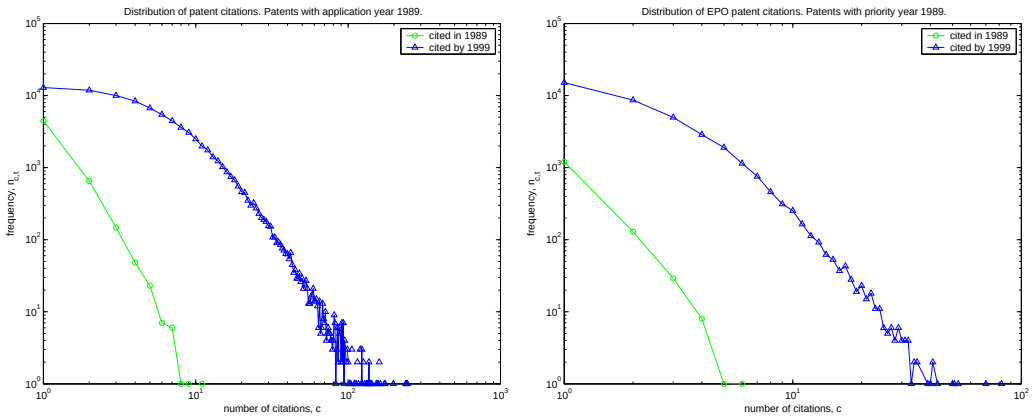

Figure 6.1: The distributions of patent citations. **Left:**USPTO cohort, application year 1989. **Right:** EPO cohort priority year 1989.

To perform the simulations and fitting we need the initial distribution of patent citations. Left diagram of Figure 6.1 also shows the distribution of the patent citations in 1989, i.e. citations within the cohort (green circles). In total there are 3,434 citations unequally distributed among the patents. Most of the patents 47,472 (94.4%) have no citation, the maximum number of citations received by a patent is 7.

I also used the data on patents granted has been collected at CESPRI. Similarly,

I limit the scope of my study to the patents with the priority year 1989. The EPO cohort of 1989 contains 61,799 patents. Due to differences in the citations practices adopted by EPO and USPTO the average number of citations per patent for the EPO patents is lower than for the USPTO patents (e.g. Breschi and Lissoni 2004), therefore for the EPO cohort of 1989 the total number of citations received is much lower than for the USPTO patents, by the end of 1999 the patents have received 99,684 citations. The most cited patent receiving 82 citations. The number of citations internal to the cohort is 1,591, with the most cited patent having received 6 citations in 1989. Both distributions of patent citations in 1989 and 1999 are shown at the right diagram of Figure 6.1.

6.5 Results

Linear feedback For a linear feedback (6.4) there is an analytical solution in the closed form (Appendix). In particular, the dynamics of the number of the patents with zero citations, n_0, is (formula (E.5) in the Appendix)

$$n_{0,t} = N\left(1 + \frac{t}{v_0 N + t_0}\right)^{-v_0},$$

where $t_0 = \sum_{i=1}^{N} c_{i,0}$, i.e. the total number of citations at $t = 0$ (citations within the cohort). The observed values of $n_{0,t}$ at t corresponding to the calendar years 1989-99 are shown at Figure 6.2. Fitting of v_0 (the only parameter in the linear model (6.4)) can be done using only values for $n_{0,t}$ (instead of fitting whole distribution). Fitting the cohort of the USPTO patents gives $v_0 = 1.1$. The results of fitting are shown at Figure 6.2.

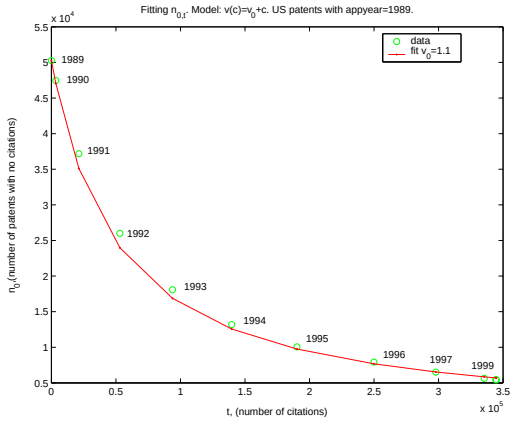

Figure 6.2: Fit $n_{0,t}$ for the USPTO cohort with linear model: $v_0 = 1.1$.

Now, with the initial distribution of patent citations and the estimated value of v_0, using equation (E.6) we can predict the distribution of the frequency of the number of citations by the end of 1999 ($t = 337,931$). The resulting distribution is shown at Figure 6.3 in linear and double logarithmic scales.[7]

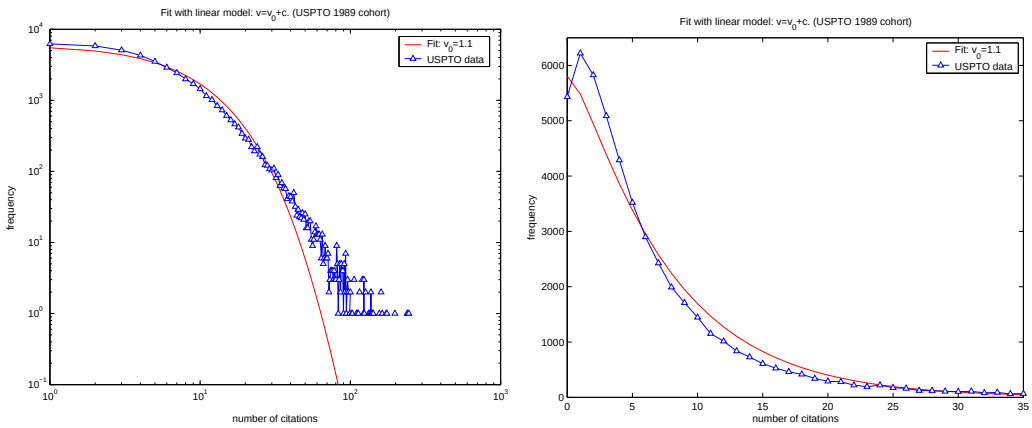

Figure 6.3: Fit of the distribution of patent citations for the USPTO cohort with linear model: $v_0 = 1.1$.

First, note that for the patents with small number of citations, the fit is good (especially if consider that we have only one parameter in the model). It indicates that the function of preferential attachment is close to linear in the region of small number of citations, c, where most of the distribution resides (98% of patents have not more than 30 citations).

Second, the tail of the actual distribution is obviously heavier than the linear model predicts (Figure 6.3). Indeed, a linear value function generates distributions with exponential tails, while the actual distribution has a Pareto-type shape for large values of c. Thus we might expect that the function of preferential attachment underlying the actual distribution of patent citations is superlinear. The nonlinearity leads to (a) effective "freezing" of the low end of the distribution at large t, because the probability that a patent with small number of citations receives additional citations is falling rapidly (faster than t^{-1}); and (b) depletion of the middle of the distribution, and as a result "fatter" tails.

Non-linear feedback The results for fitting the distribution of USPTO patent citations with simulated distribution, in case of non-linear feedback in the form (6.5) are shown at the top diagrams of Figure 6.4. The values of parameters providing the best fit are $v_0 = 2.0$ and $\alpha = 1.26$. As one can see the simulated distribution

[7]Fitting the distribution of the EPO citations produces similar results and is not reported here.

fits observed distribution very well for most values of c (there is an overshooting at $c = 1$).

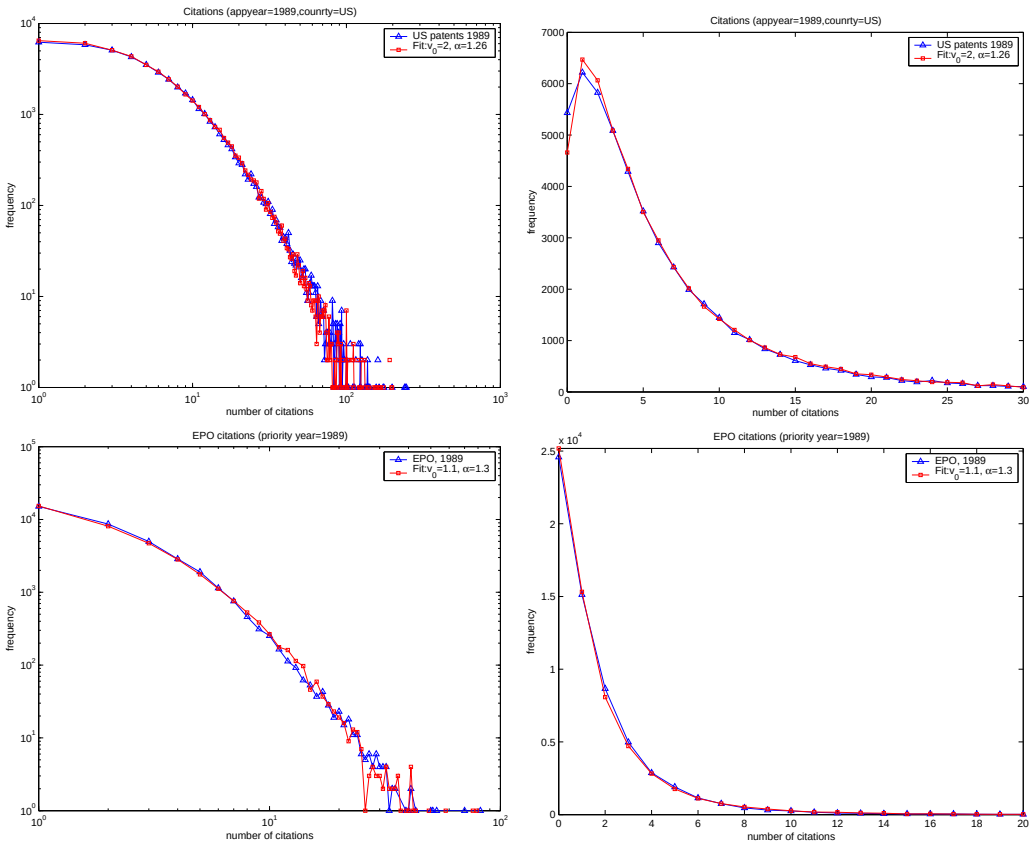

Figure 6.4: Fit of the distribution with non-linear model. **Top:** USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Bottom:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.

Fitting the EPO data using the same procedure gives $v_0 = 1.1$ and $\alpha = 1.3$. A fit (in linear and double logarithmic scale) is shown at the two bottom diagrams of Figure 6.4. The lower "propensity to cite" of EPO patents mentioned above reveal itself in the lower value of parameter v_0. However, and more important, the value of the parameter α describing non-linearity and controlling the shape of the middle range and the tail of the distribution is not that different from the value of α providing the best fit for the USPTO cohort.[8]

Figure 6.5 shows quantile-quantile plots (QQ-plots) for the simulated distributions vs. observed distributions. If the data falls on 45^0 line of a QQ plot, it means

[8]Experiments with fitting the whole 1989 cohort of the USPTO patents without selection on the country of the first inventor (96,077 patents) and the cohort of the USPTO patents applied for in 1975 (with citations received from 1975 to 1999) give slightly different values of v_0, but rather robust on the value of parameter $\alpha \approx 1.2$–1.3.

that the distributions underlying the samples of observed and simulated are identical. As one can see from the Figure 6.5 the quantiles of the simulated distribution for the EPO cohort are lying on the 45^0 line until approximately 40 citations, which is 99.99-percentile of the observed sample (30 citations is the 99.96 percentile). For the USPTO cohort reasonable fit is achieved from 0 to about 150 citations which includes 99.98% of patents (100 citations correspond to 99.92 percentile).

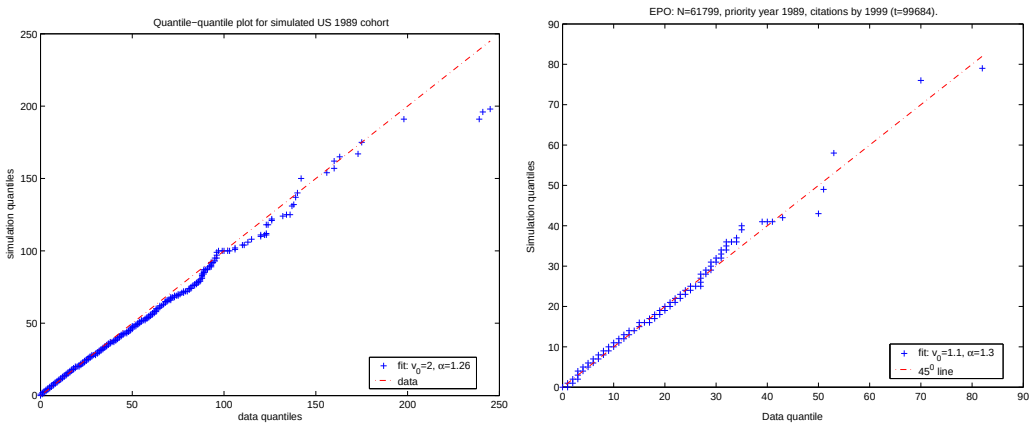

Figure 6.5: QQ-plots for the simulated distributions, non-linear model. **Left:**USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Right:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.

The simulated distributions have fat tails. The tail index of the distribution (the exponent in the Pareto distribution describing the tail) can be estimated using the Hill estimator (Hill 1975)

$$\gamma_{N,k} = \frac{1}{k}\sum_{i=1}^{k}(\ln c_{(i)} - \ln c_{(k+1)}),$$

where $c_{(1)} \geq c_{(2)} \geq \cdots \geq c_{(N)}$ denote order statistics. A Hill plot, the diagram of the inverse of the Hill estimator, $1/\gamma_{N,k}$, vs. the rank of the observation, k, can be used to learn about the tail index and the cut-off value of the tail: the value of $1/\gamma_{N,k}$ at which the plot stabilizes provides an estimate for a tail index, and the value of the corresponding order statistic gives the cut-off value for the tail. The Hill plots for observed and simulated data for the USPTO and the EPO cohorts are shown at Figure 6.6 (left: USPTO data $v_0 = 2.0$, $\alpha = 1.26$, right: EPO data $v_0 = 1.1$, $\alpha = 1.3$). The plot stabilizes at value of α somewhere between 3.0 and 4.0 for both cohorts, i.e. the exponent in the Pareto distribution exceeds 2.0 therefore the distribution has finite mean and variance.

It is interesting to compare our results with the results of Trajtenberg (1990) who used value functions similar to (6.2) and (6.3) (with fixed $v_0 = 1$) in the study of patents in CT scanner technology for construction of weighted patent counts (WPC).

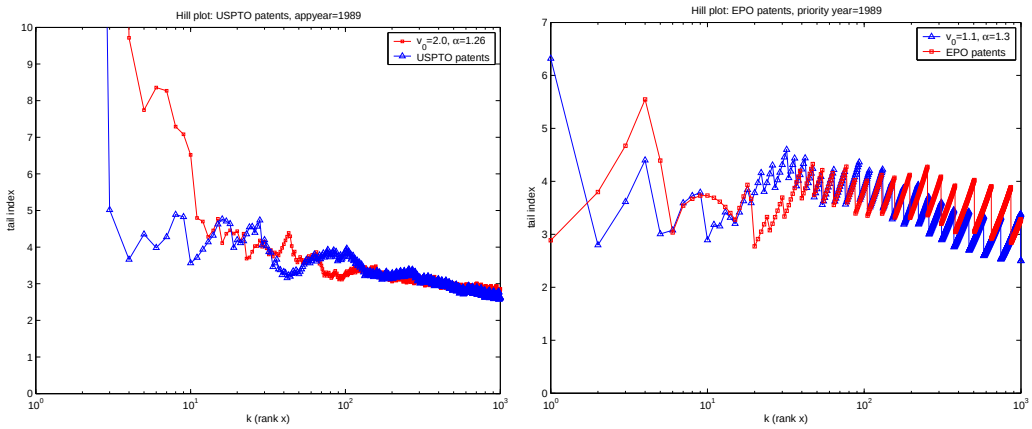

Figure 6.6: Hill plot for the observed and the simulated distributions, non-linear model. **Left:**USPTO cohort $v_0 = 2.0$, $\alpha = 1.26$. **Right:** EPO cohort $v_0 = 1.1$, $\alpha = 1.3$.

He found that WPC have significant (cross-time) correlation with the social value of innovations estimated via demand for the new models of scanners. For the non-linear WPC the best results were obtained with $\alpha = 1.3$ and $\alpha = 1.5$. Although our data and approach are rather different, the value of the parameter α providing the best fit falls in the same range.[9]

It is also worth mentioning that Hall, Jaffe, and Trajtenberg (2005b) in their study of the impact of company's stock of patents on the market valuation of the company found that the relationship between the market valuation and the number of citations received by the patents owned by the company is non-linear - while the impact of citations is not significant for patents with low number of citations, the magnitude of the effect becomes significant as the number of citations grows. A comparison between the results of the simulations with the linear (6.4) and non-linear (6.4) models lead to the conclusion that the value function $v(c)$ is non-linear.

6.6 Discussion

Let us turn to limitations and possible extensions of the model. First, I would like to elaborate on the problem of "intrinsic values" of inventions and their relationship with the productivity gain. Then I will make a brief remark on the omission of the variation mechanism. At the end of the section I will discuss the use of technological fields conveyed by patent classification to describe the path-dependent process of technical change.

Many models concerning technical change emphasise that the main characteristic

[9] According to (Trajtenberg 1990) the difference in correlation between $\alpha = 1.3$ and $\alpha = 1.2$ are only several percentage points.

of an innovation is the increase of the productivity which this given innovation offer once it is adopted. From this perspective the "intrinsic value" of an innovation is already predetermined and mostly (if not solely) depends on the productivity gain which is assumed to be distributed according to some probability law. Therefore, there is no question about the process that govern the dynamics of the values, but an inquiry about the distribution of patent values can be safely reduced to the question about the exact form of the distribution of the productivity gains. This view is in sharp contrast with the model proposed in this paper. Indeed, the model assumes that all innovations are "born" equal, and it is selection which following the evolutionary theory of technical change that generates the differences in the values.

Surely, the value of an innovation depends on many factors besides the direction of technical change, including productivity gain, but also demand for the new product, advances in science and so on. Acknowledging the importance of factors other than the productivity gain, I shall remark on the latter, primarily because as has already been mentioned, most models take it as a premise.

It is certainly true, that if we consider a range of alternative technologies which were developed some time in the past to address a certain design problem, then a technology dominating the market at present is more likely to be more efficient. However, to conclude that at the time of invention it had higher "intrinsic value" related to its efficiency in comparison with other alternatives might be an unjustified stretch.

From the history of technology we know many examples when with respect to productivity a newly born technology had been inferior to the existing one and only incremental improvements over a long period of time let these technologies prevail.[10] The reason why innovators spent their time on working with seemingly inferior technologies is that these technologies, while being less productive, offered a basis for a technological breakthrough, and reasons why these technologies have surpassed the alternative designs are rooted in the complementarities between technologies (Rosenberg 1982).

Let me illustrate this point with the results from the percolation model of Silverberg and Verspagen (2005) mentioned in Section 6.2. Consider the technology space in a form of two-dimensional lattice, with the vertical dimension representing productivity (with more productive technologies at the top and less productive ones in the lower part of the lattice). A technology becomes available only when at least one adjacent technology is already in use. Initially agents know only technologies at the bottom. Growth in such a model is the process of percolation from the bottom to the top of the lattice. If all technologies had the same probability to be discovered, growth would occur along a line(s) connecting bottom and top of the lattice. However, linear growth is prevented by a random "landscape": each point of the

[10] We can only guess how many potentially valuable technologies never made it through.

lattice representing a certain technology has different probability to be discovered. Consider the extreme case when the probability of discovering technology which is the next on the "linear expansion path" is zero. If the search were constrained to the area of the technology space just above the most efficient current technology, then the technological progress would cease forever. Nevertheless, it proceeds due to the agents who keep searching in areas of less productive technologies which at the end results in finding a "side-path". What is important in our context is that most productive technologies does not have to be the most promising, and once the growth is stuck it is less productive technology that can make a difference, if it can lead out of the "deadlock".

Therefore judgments about "intrinsic value" *ex-post*, conditional on the success or failure of technologies might oversimplify the complex picture of technical change. The *ex-post* value as measured by patent citations is the result of a path-dependent process and reflects different factors such as productivity gains, demand conditions, complementarity with other innovations, and some (mis)fortune. It hardly can be reduced to the productivity gain alone.

Having stated that, I nevertheless shall note that the approach presented in this paper can be (and should be) improved. Emphasising importance of path-dependency in the evolution of innovation values, I have omitted the fact that the innovations in consideration (patent from the 1989 cohorts in our case) were not born in vacuum, but also were a consequent development of some earlier technologies. Once we assume that the current value of an innovation depends on how well the innovation is embedded in the current "technological paradigm" reflected by the number of "forward citations", we can make one step further and assume that the initial value of an innovation i, $v_{i,0}$, depends on how well the innovation was embedded in the paradigm at the time of the invention, and hypothesize that "backward citations" i.e. citations made by the patent convey some information about it.[11]

Another limitation of the study presented here is that focusing on the patents from one cohort I have restricted the scope of the analysis to the *selection mechanism*, omitting the other main component of the evolutionary process - the *variation mechanism*, the mechanism that generates new technologies and leads to the discontinuities in the technological trajectories. Some features of the variation can be traced in the patent data. For example, "aging" of patents, i.e. the decline in the rate of receiving citations with time may be a reflection of the shifts in technological trajectories in different subfields of the technology. Moreover, the process of formation of new technologies might be reflected in the patent classification, the point to which we will come late in this section. However, to conduct the study of variation mechanism based on patent citations, one has to find a solution for the problem of

[11] Deng et al. (1999) report that the number of backward citations is indicative for the value of innovations.

"benchmarking" mentioned in the Section 6.2, because such a study cannot be done without inter-cohort comparisons.

There is also a problem related to the fact that patents which we consider do not belong to the same technological field. It raises two issues. First, it is well known that different industries have different "propensity to patent". Therefore, it may be that the actual distribution of patent citations is sheer reflection of this fact rather than a result of the path-dependent process similar to one proposed in this paper. Underlying this question is a suspicion that if we restrict our analysis to one technology then the shape of the distribution of patent citations may be quite different from the shape of the overall distribution. On the other hand, if the model is correct, then at the level of a patent class (or related patent classes) we expect to see the distribution of patent citations similar to one on the level of the whole cohort.

First, note that there are, indeed, differences in the average number of citation in different classes that can be attributed to differences in the "propensity to patent" across industries: distributions of patent citations from different patent classes occupies different range of distribution. For example, the USPTO patents related to data processing (USPTO patent classes 700-714) are on average more heavily cited through 1989-1999. However, inspection of the distribution of the patent citations within the same patent class (or related patent classes) reveals the picture similar to one we have seen at the level of the whole cohort.

Figure 6.7 shows the distribution of patent citations for USPTO patents related to data processing (classes 700-714) applied in 1989. As one can see it is strikingly resembles distribution of patent citations for the whole cohort of 1989 (Figure 6.1): it is highly skewed and has a long tail. The best fit is obtained with parameters $v_0 = 4.0$ and $\alpha = 1.26$. The value $v_0 = 4.0$ is twice as high as the fitting value for the whole cohort ($v_0 = 2.0$), which reflects the fact that through 1989-99 the patents in data processing have been cited more frequently than patents from other classes. However the parameter $\alpha = 1.26$ controlling the shape of the centre and top end of the distribution is the same as for the whole population, which implies that the functional form of the value function (except the shift of intercept) is the same as for the whole cohort.

The second issue concerning the technological field is related to the boundaries within which a company can reallocate its R&D activities responding to shifts in technological trajectories. The model assumes that exploring opportunities opened by previous inventions a company chooses to search in the area of the technology space that is "popular". It does not contradict economic intuition, when the reallocation is to take place within the same technological field, however if it is to be done across industries, then, at least, one need an explanation: afterall, why a company producing, say, domestic appliance should be investing in nanotechnology?

There are two reasons which could partially justify assumptions of the model.

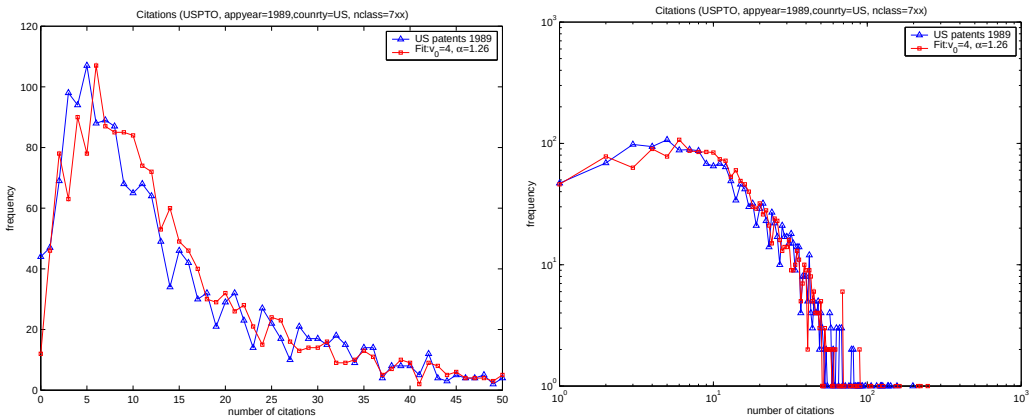

Figure 6.7: The distribution of patent citations for patent classes 700-714 (Data processing) with application year 1989, and the USA as a country of the first inventor.

First, most patents is assigned to large diversified companies (such as IBM), or industrial conglomerates involved in innovative activities in many R&D intensive sectors, and reallocation of innovative activities by such company corresponds to the reallocation of R&D budgets within a company or a conglomerate. Second, highly cited patents are likely to be related to General Purpose Technologies (GPTs) (Hall and Trajtenberg 2004), the technologies which penetrate most sectors of the economy. Therefore companies in different sectors may be involved in adapting a GPT to their needs, and it is reflected in the observed pattern of patent citations.

However reasonable it seems, the model needs to be modified to take into account the fact that many economic entities performing R&D are specialized in certain sectors. For example, we may change the model in such a way that the selection of the technology from which to start the R&D search, and as a result which patent will be cited is done in two steps. At the first step, the sector in which a new patent is to be taken will be selected, and then a particular technology (represented by the corresponding patent), which is to be used as a starting point will be chosen on the basis of the values of technologies in this field.

This modification of the model, in turn, opens a question of how to choose the technological field for a new patent. For that we can use information about the patent classes (as a representation of separate fields). The problems concerning the use of patent classifications are discussed below. If the selection of the technological field is done in the way similar to the one which we use in our model of patent citations, i.e. the probability of a patent to appear in a certain sector is a function of the number of patents in this field in comparison to the whole stock of patents in all patent classes, then we would have some kind of a "nested" Polya process.[12]

[12]If both value functions at both stages of selection are linear then the two-stage process is

Our model states that the more R&D have been done in a certain field (resulting in more patented inventions), the more R&D effort will be directed to this field in close future. Translating the assumptions of the model into the context of patent classes we would expect that the larger is the share of a patent class in the stock of patents, the higher is the probability that the next patent will appear in this patent class. [13] To check if this intuition is correct, in Figure 6.8 (left diagram) I plot the share of a USPTO patent class (417 patent classes) in the stock of the patents applied in each of the years 1989-1999 ($n_{i,t}/n_t$) against the share of the patent class in the stock of the patents from 1963 to the respective year ($N_{i,t}/N_t$). Take for example, 932 patents applied in 1989 in the patent class 29 "Metal working"($n_{29,1989} = 932$). In 1989 the number of patent applications to all classes, n_{1989}, was 96,077, it gives us the share of the class 29 in the stock of all patents applied in 1989, $n_{29,1989}/n_{1989} = 932/96,077 \approx 0.0097$. Now, from 1963 to (not including) 1989 there were 20,323 patent applications in the class 29 ($N_{29,1989} = 20,323$). The total stock of all patents from 1963 to 1989 is $N_{1989} = 1,878,708$, therefore the share of the class 29in the total stock of patent applications in 1989-1999 is $N_{29,1989}/N_{1989} = 20,323/1,878,708 \approx 0.0108$. As one can see from Figure 6.8 the observations reside close to 45^0 line.[14]

Figure 6.8 (diagram on the right) shows the evolution of several patent classes (circles mark points in 1989). Generally we can divide all patent classes into three broad categories according to their growth patterns: mature technologies with stable shares, old technologies with shrinking shares, and new technologies with growing shares. As one can see from Figure 6.8 patent class 29 "Metal Working" containing 29,858 patents, or 1.02% of all patents from 1963-1999, has a stable share in the total patent stock, and the rate of arrival of new patented inventions in this class is proportional to its share. Classes 435 ("Chemistry: Molecular Biology and Microbiology", 30,257 patents or 1.03%), 436 ("Chemistry: Analytical and Immunological Testing", 6,998 patents or 0.24), and 514 ("Drug, Bio-Affecting and Body Treating Compositions", 58,062 patents or 1.99%) are rapidly expanding through 1963-1999. At the same time the shares of class 12 ("Boot and Shoe Making", 1,251 patents or 0.04%) and class 66 ("Textiles: Knitting", 3,846 or 0.13%) are going down. These developments in patenting activities are not necessarily related to the current shares of the corresponding industries in the total output, but we might expect that they reflect long-term trends in the economy.

The distribution of the USPTO patents in the patent classes and the results of fitting with lognormal and gamma distributions is shown at Figure 6.9. The goodness-of-fit statistics are reported in Table 6.1. Although goodness-of-fit for

observationally equivalent (i.e. distribution of citations is the same) to one in simple one-stage citations model (6.4).

[13]This maps exactly into the model but at a higher level of aggregation.

[14]The slope in double logarithmic scale slightly exceeds 1.0 indicating a superlinear relationship between the variables, akin to the non-linear model discussed earlier.

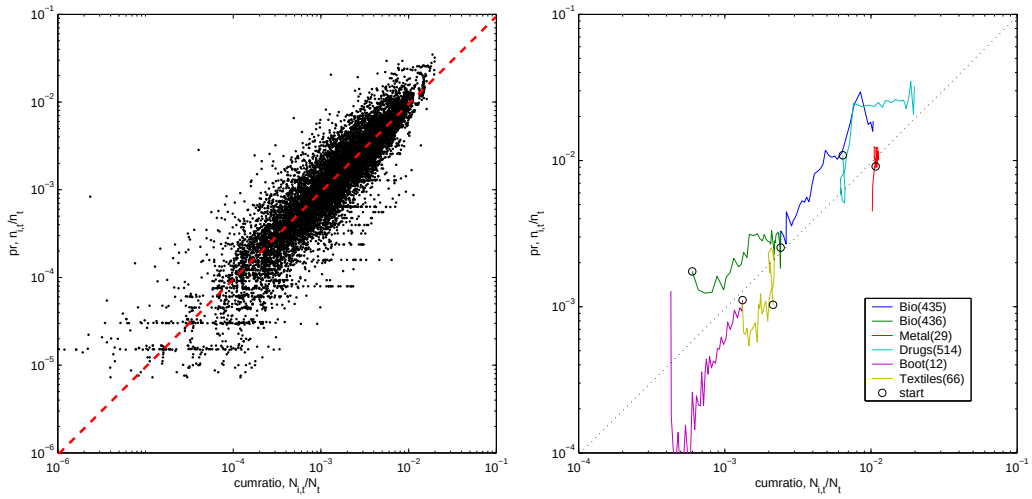

Figure 6.8: **Left:** Share of a USPTO patent class in the stock of patents applied in year t (from 1990 to 1999), $n_{i,t}/n_t$ vs. the share of a patent class in the whole stock of patents applied since 1963,$N_{i,t}/N_t$. **Right:** Diagram $n_{i,t}/n_t$ vs. $N_{i,t}/N_t$ for several USPTO patent classes. Circles mark positions in 1989. Classes: 29-Metal working, 435-Chemistry: Molecular Biology and Microbiology, 436-Chemistry: Analytical and Immunological Testing, 514-Drug, Bio-Affecting and Body Treating Compositions, 12-Boot and Shoe Making, 66-Textiles: Knitting.

both lognormal and gamma distributions are reasonable ($p < 0.01$), the Gamma distribution is marginally better. Notice, that formula (E.4) derived for the finite Polya process with linear feedback predicts the distribution close to gamma distribution.[15].

Coming back to our research question, this information could be used for building a two-stage model as described above. However, there are also some difficulties here related to patent classification. First, there is inherent ambiguity to which industry (and related patent class) an invention should be assigned. An invention may be assigned to a class on the basis of the industry from which it originated, the industry that will produce the new product, or the industry which will use it (Griliches 1990). As a result, developments of the same technology may be divided among different patent classes. Another problem, also related to the interconnections between patent classes, is that a patent class hardly can represent a whole industry or a sector. Therefore to proceed with a two-stage model one has to decide how to aggregate classes into industries.[16]

Notice also, that the patent classification is evolving with the technology: new patent (sub)classes are being added, reclassified etc. Relying on the current classification one necessarily has some kind of bias when making judgement about past

[15]To see this one can use the representation of Beta function through Gamma functions

[16]Several ways to do it are outlined in (Hall and Trajtenberg 2004).

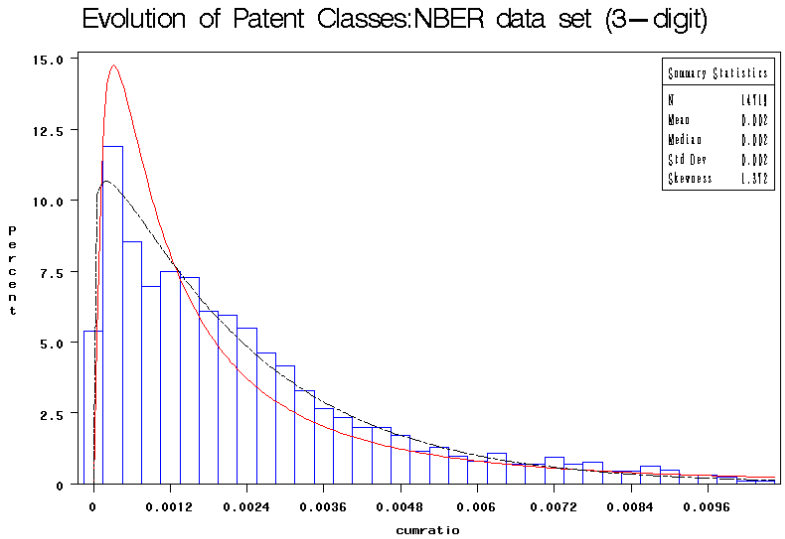

Figure 6.9: The distribution of USPTO patent class sizes. Patents granted from 1963-1999. Fit: red solid line - LogNormal, black dash-dot line - Gamma.

Table 6.1: Goodness-of-Fit Tests for the Distribution of Patent Class Sizes.

Test (Statistic)		Lognormal		Gamma	
		Statistic	p Value	Statistic	p Value
Kolmogorov-Smirnov	(D)	0.091486	< 0.010	0.0234421	< 0.001
Cramer-von Mises	(W-Sq)	35.711508	< 0.005	1.6382987	< 0.001
Anderson-Darling	(A-Sq)	204.837144	< 0.005	10.8396215	< 0.001

inventions. For example, if one is to use current classification on some fine level, say, 6-digit subclasses, then one might be surprised by discovering that a number of subclasses were unpopulated back in the 70s. These subclasses have been added as the corresponding technology came into being. In terms of the model the situation with addition of patent (sub)classes should be modelled with an infinite Polya process, where new bins are constantly being added. It is also possible to use more general class of the processes, Yule processes (Yule 1925), earlier applied in evolutionary biology. An advantage of the Yule process is that it not only accounts for the addition of the new classes, but also traces the lineage of the evolutionary tree.[17]

[17]For a description of the Yule processes and the properties of the distributions generated in these processes see (Newman 2005).

6.7 Conclusions

Taking the prospective of the evolutionary theory of technical change I argued that the value of innovations depends on the direction of technical change, and patent citations reflect the path-dependencies in the development of technology. Innovations well embedded in the current "technological paradigm" have higher value and play a role of "focusing devices" shaping direction of innovation search. As a result the higher is the value of an innovation the more likely it is to be used as a starting point for consequent innovations and the more valuable it will become. To formalize this argument I proposed a simple model based on generalized Polya processes with linear and non-linear (positive) feedback.

Using NBER and CESPRI data on patents granted by the USPTO and EPO I have shown that the model does produce distribution of patent citations close to the observed one. The model with a linear feedback predicts correctly the distribution of patent citation for patents with relatively small number of citations (about 95% of the distribution). The model with non-linear feedback predicts the distribution of patent citations correctly for the whole range of citations. Simulated distributions do have fat tails (as the observed distributions). Interestingly, the exponent in the feedback function providing the best fit is in the same range as the estimate of Trajtenberg (1990) obtained in a different context.

Chapter 7

Conclusions

Social learning and imitation are essential parts of a powerful mechanism through that a novel behavioural pattern is transmitted through the society. In **Chapter 2** we have surveyed the body of research done in a range of behavioural sciences on the social learning and imitation. We concluded that imitation is not peculiar to human beings, but rather common among other animals. However for our species the survival in the course of evolution has been primarily dependent on the (high) propensity for imitation and social learning. Therefore, at least partly, human tendency to copy others behaviour is embedded in our genes. Organization theory literature extends the notion of imitation to describe the behaviour of organizations. According to the mainstream economics the motivation for copying behaviour of others traditionally has been assigned to direct or indirect payoff externalities. More recently economists turn their attention to boundedly rational behaviour and developed a number of models to analyse properties of an economic system where agents imitate behaviour of each other. We have also examined effects of social learning in some particular settings characteristic to economic environment to provide the background for the economic contexts which we explored in the following chapters.

As the ICT revolution enhances density of communication network, on the one hand, and the increase the geographical reach of information flows, on the other, to understand better the effects of such ICT-based globalization in **Chapter 3** we inquer in the relationship between the structure of information flows and efficiency of social learning. To address this problem we develop and analyse a model of social learning in networks following the lines of Bala and Goyal (1998,2001).

First, we noticed that according to BG's model of social experimentation the increase in the "degree of integration" of the society that we can relate to the effect of the ICT revolution unambiguously enhances efficiency of social learning, as it reduces the probabilities of pathologies in the configurations of the network related to insufficient connectivity of the society (such as presence of a "royal family") and having negative effects on the social welfare.

We asked whether this holds true when (a) the scope of publicly disclosed information is limited, in particular, when the outcomes of the agents' actions are kept private and only the actions are observable, and (b) agents' decisions are irreversible. To answer this question we formulated and examined a model of informational cascades Bikhchandani, Hirshleifer, and Welch (1992), where similarly to BG's we assumed that agents may observe only their closest neighbours. To disentangle effects of the "death of distance" and increasing density of the communication network we examined informational cascades in the networks generated by "β-algorithm" of Watts (1999).

Our results suggest that, first, the higher is the share of "global contacts" between distant communities, the higher is the social welfare (in terms of the average payoff). Second, similarly to Bala and Goyal the higher the "degree of integration" within the society is, the more likely it is that conformity of actions will arise. However, unlike their results our model suggests that in the presence of informational externalities globalisation of informational flows, expressed in the increasing density of communication channels in a network, may drive down the expected social welfare.

In many economic environments there are distinct types of the agents playing distinct economic roles (e.g. suppliers and consumers, companies and their investors) and an economic transaction depends on the joint action of actors of different types. It is also often the case that the information available to the agents of different types is different. In **Chapter 4** we examined the process of social learning in such an environment. We argued that social learning in an environment with two types of the agents on the two sides of the market asymmetrically distributed information may lead to "mutual illusions". We devised a model in which agents on the two sides of the market are subject to informational cascades, and find that in an uncertain environment with asymmetric information agents tend to be overoptimistic about the state of the world, a result which fits with empirical evidence on financing new technologies. This overoptimism based on mutual illusions makes the system vulnerable to two-sided bubbles, and may be one of the reasons behind "dot com" crash.

Individual ability to imitate is not specific to human beings, but is an intrinsic feature of all social animals. However emulation as a *social process* in human societies in many respects is strikingly different emulation in animal world, as, on the one hand, no other species match humans with respect to complexity and sophistication of social organization, and on the other hand, the social structure and the institutions on which this structure is built have their mark on all social processes. The process of emulation is no exception.

In **Chapter 5** we discussed how two particular characteristics of the society, class structure and social norms related to conspicuous consumption, shape the process of product innovation. Using a simple evolutionary model of diffusion of a

positional good we have shown that a society with more equal class structure and social norms encouraging behavioural variety is characterized by both high market penetration and high speed of diffusion. We related those results to the debate on the importance of demand factors in the Industrial revolution, and argue that the changes in the social structure and social norms regarding consumption prior to the industrialization in western Europe prepared the ground for introduction of the industrial methods of production.

In the evolutionary theory the path-dependency in the process of technical change is often attributed to the existence of "focusing devices / technological guideposts / technological trajectories", i.e. the theory emphasises the role of the cognitive factors framing the direction of innovation search. From our perspective we can easily recognize the elements of the social learning behind those factors. In **Chapter 6** we examined how the framework of path-dependency and technological trajectories can be applied to explain the observed distribution of patent values revealed in the distribution of patent citations. A very simple model of based on generalized Polya urn processes has been proposed to explain the evolution of patent values. We have confronted the results of the model with the observed distributions of patent citations for two cohorts of patents from two major patent offices, USPTO and EPO. Surprisingly, this simple model fits empirical distribution of patent citations (USPTO and EPO data) very well. We found that the relationship between the number of citation received by a patent ("forward citations") and the social value of the underlying innovation is likely to be non-linear. The results of fitting the observed distributions with our model suggest that the values of the exponents describing non-linearity for the two cohorts lie in the same range, despite the differences in the citation practices between USPTO and EPO.

Appendix A

BHW model

Decision rule. Consider the decision problem of an agent who is to take his decision, $\Delta_i \in \{Adopt, Reject\}$, at time t. Information available for him, I_t, consists of his private signal $v_t \in \{High, Low\}$, and the history of agents' actions prior to t, H_t. Expected payoff to adopting is

$$E[u_t] = 1 \cdot \Pr(V = 1|I_t) + (-1) \cdot \Pr(V = -1|I_t).$$

He adopts if $E[u_t] > 0$, rejects if $E[u_t] < 0$. In case of a 'draw', $E[u_t] = 0$, he follows his signal, i.e. adopts provided that $v_t = High$, otherwise he rejects.

Applying Bayes' rule and taking into account that signals are conditionally independent (therefore $\Pr(I_t|V) = \Pr(v_t|V)\Pr(H_t|V)$) we can find the condition for an UP cascade, the situation where he adopts regardless to the signal (priors $\Pr(V = 1) = \Pr(V = -1) = 1/2$)

$$(1 - p)\Pr(H_t|V = 1) > p\Pr(H_t|V = -1).$$

Similarly for a DOWN cascade

$$p\Pr(H_t|V = 1) < (1 - p)\Pr(H_t|V = -1).$$

Suppose that by time $(t - 1)$ no cascade has happened, hence agents' actions unambiguously reveal their signals. Let n_+ be the number of adopters, and n_- be the number of non-adopters. Then the condition for UP cascade can be rewritten as

$$p^{n_+}(1 - p)^{n_- + 1} > (1 - p)^{n_+} p^{n_- + 1} \Leftrightarrow \left(\frac{p}{1 - p}\right)^{n_+ - n_- - 1} > 1.$$

By asssumption $p > 1/2$, therefore condition for UP cascade is

$$n_+ - n_- - 1 \geq 1 \Leftrightarrow n_+ - n_- \geq 2.$$

Similarly for a DOWN cascde

$$p^{n_+ +1}(1-p)^{n_-} > (1-p)^{n_+ +1}p^{n_-} \Leftrightarrow \left(\frac{p}{1-p}\right)^{n_+ - n_- +1} < 1.$$

or

$$n_+ - n_- \leq -2.$$

Thus the decision rule is: if by the time when the agent is to take the decision the difference between the number of the adopters and the number of the non- adopters, $d \equiv n_+ - n_-$, is greater or equal to +2, then he adopts regardless of the private signal; if the difference is less or equal to -2, the agent rejects whatever is the private signal; if neither is the case, he chooses to follow the signal, i.e. to adopt if the signal is $High$ and reject when it is Low.

Shares of cascades, total outcome. Without loss of generality let us assume that $V = 1$. The decision rule described above implies that an informational cascade may emerge only when t is odd. Suppose that there is no cascade by date $t = 2(l-1)$. An UP cascade will emerge at $t = 2l + 1$, if both agents at $t = 2l - 1$, and $t = 2l$ receive positive signals. Therefore, the probability of UP cascade is p^2. Similarly, the probability of a DOWN cascade is $(1-p)^2$, and the probability of no cascade is $q \equiv 2p(1-p)$.

If an UP cascade emerges at $t = 2l+1$, then the share of agents in the UP cascade is $(n-2l)/n$. Therefore the share of the agents to be locked in an UP cascade is

$$S_{up} = \sum_{l=1}^{K} \frac{2K-2l}{2K} p^2 q^l = \frac{p^2}{1-q}\left(1 - \frac{q}{(1-q)K} + \frac{q^{K+1}}{(1-q)K}\right), \tag{A.1}$$

where $K = n/2$. Note, that the factor $\frac{p^2}{1-q}$ is the probability that the system eventually ends up in an UP cascade in the limit of $n = \infty$, P_{up}. It is rather natural to expect: the system converges to one of the cascades very fast (exponenially) therefore the share of agents which follow their signals is decreasing ($\sim n^{-1}$) and S_{up} must aproach P_{up}. Similarly, for DOWN cascades

$$S_{down} = \sum_{l=1}^{K} \frac{2K-2l}{2K} (1-p)^2 q^l = \frac{(1-p)^2}{1-q}\left(1 - \frac{q}{(1-q)K} + \frac{q^{K+1}}{(1-q)K}\right). \tag{A.2}$$

The probability of DOWN cascade in infinite population P_{down} is the asymptotic value for S_{down} for large n.

If an UP cascade emerges at $t = 2l + 1$, then the total outcome normalized by n is $(n - l + 1))/n$, and if at this time a DOWN cascades emerges then the total

outcome is $(l-1)/n$. Therefore

$$W = \sum_{l=0}^{K-1} \frac{(2K-l)p^2 + l(1-p)^2}{2K} q^l = \frac{1}{2} + \frac{2p-1}{2(1-q)} \left(1 - \frac{q}{(1-q)K} + \frac{q^{K+1}}{(1-q)K}\right). \tag{A.3}$$

Shares of cascades, S_{up} and S_{up}, total outcome, W, as functions of the size of population $(n = k+1)$ are shown at Figure 3.2 $(p = 0.75)$.

Appendix B

Mean-field solution for network cascades

Consider an agent i who is to make his choice, $\Delta_i \in \{Adopt, Reject\}$, at time t. His decision depends on three (random) variables: the number of adopters, ξ_i, and the number of non-adopters (those who rejected the technology), ζ_i, in ith neighbourhood Γ_i, and the private signal $\sigma_i \in \{High, Low\}$, in the following way

$$\Delta_i = \begin{cases} Reject, & \text{if } (\xi_i - \zeta_i) \leq -2, \\ Reject, & \text{if } |\xi_i - \zeta_i| < 2 \text{ and } \sigma_i = Low, \\ Adopt, & \text{if } |\xi_i - \zeta_i| < 2 \text{ and } \sigma_i = High, \\ Adopt, & \text{if } (\xi_i - \zeta_i) \geq 2. \end{cases} \tag{B.1}$$

Private signals σ^t, $t = 1..N$ are drawn at random with probabilities $\Pr(\sigma^t = High|V = 1) = p > 1/2$ and $\Pr(\sigma^t = High|V = -1) = 1 - p$. Values of ξ_i and ζ_i are determined by the history of signals $\{\sigma^1, \sigma^2, .., \sigma^{t-1}\}$ and the structure of the network G.

The state of the system is described by vector $S^t = (s_1^t, s_2^t, .., s_N^t)$, where s_i^t is the state of agent i: s_i^t is equal to Δ_i, if he already made his decision by time t, otherwise $s_i^t = 0$. The evolution of S^t is determined by individual decisions according to (B.1) and the structure of G. To solve the dynamics means to find the distribution of S^t from the distribution of the realizations of the signals $\{\sigma^1, \sigma^2, .., \sigma^{t-1}\}$ and the distribution of the order in which the agents make their decisions. The problem already seems to be complicated, but to make the bad things worse we have to take into consideration that for $\beta > 0$ the small-worlds algorithm generates G randomly and to answer the questions we are interested in, we should not only find the distribution of statistics of our interest (S_{up},S_{down},and W) at the terminal time $t = N$, but also obtain the distribution of those statistics over the 'ensamble' of the small-world networks generated for certain value of rewiring parameter β.

Although the problem is rather complicated, we can relatively easily find an approximation of the dynamics using mean-field approach. First, we approximate the random variables by their averages, hence before the emergence of a cascade the share of adopters and non-adopters are exactly p and $(1-p)$ respectively, i.e. we neglect fluctuations in realizations of the private signals. Second (and more brutal), instead of analysing what happens in agents' i (idiosyncratic) neighbourhood we are going to analyze a representative neighbourhood $\bar{\Gamma}$ of a representative agent. The term 'representative' in this context stands for 'averaged over the population', i.e. this assumption rules out all possible local fluctuations. This is a very rough approximation of the network G, justified only if local fluctuations do not grow. Nevertheless for networks such as trees where no i's neighbours are neighbours of each other (so that informational cascades cannot spread), it may produce a reasonable approximation.

Let N_a^t be the number of adopters by time t, and $N_{na}^t = t - N_a^t$ the number of non- adopters (once we leveled down the description of the structure of G to representative neighbourhood, the state of the system can be fully described by these two numbers). Then the number of adopters in $\bar{\Gamma}$ is $k\frac{N_a^t}{N}$, where k is the average degree of a node in G (approximation for ξ_i). Similarly, $k\frac{t-N_a^t}{N}$ is the number of non-adopters in $\bar{\Gamma}$ (approximation for ζ_i). The difference between the number of adopters and non-adopters in $\bar{\Gamma}$ is

$$\bar{d}^t = k\frac{N_a^t}{N} - k\frac{t-N_a^t}{N} = \frac{k}{N}(2N_a^t - t). \tag{B.2}$$

Therefore, mean-field approximation for evolution of our system can be written as (compare with (B.1))

$$N_a^t = \begin{cases} N_a^{t-1}, & \text{if } \bar{d}^t \leq -2, \\ pt, & \text{if } |\bar{d}^t| < 2, \\ N_a^{t-1}+1, & \text{if } \bar{d}^t \geq 2; \end{cases} \quad \text{and } N_{na}^t = \begin{cases} N_{na}^{t-1}+1, & \text{if } \bar{d}^t \leq -2, \\ (1-p)t, & \text{if } |\bar{d}^t| < 2, \\ N_{na}^{t-1}, & \text{if } \bar{d}^t \geq 2. \end{cases} \tag{B.3}$$

By assumption $p > 1/2$, hence there no DOWN cascades in the mean-field solution. An UP cascade emerges if N is not too small. More precisely, UP cascade emerges at t^* as N_a reaches the 'critical value', N_a^*, such that $\bar{d}^t = 2$. From (B.2) and (B.1) we can find the time when UP cascades starts (if it does)

$$\begin{cases} \frac{k}{N}(2N_a^* - t^*) = 2, \\ N_a^* = pt^*; \end{cases}$$

or

$$t^* = \tfrac{2N}{k(2p-1)}, \quad \text{if } k(2p-1) \geq 2. \tag{B.4}$$

The condition in (B.4) insures that $t^* \leq N$.

From (B.4) the share of agents who are locked in an UP cascades, S_{up}, is

$$S_{up} = 1 - \begin{cases} 0, & \text{if } k(2p-1) < 2, \\ 1 - \frac{2}{(2p-1)k}, & \text{otherwise.} \end{cases} \tag{B.5}$$

Prior to t^* the number of non-adopters grows at rate of $(1-p)$: $N_{na}^t = (1-p)t$; at t^* the growth ceases and N_{na}^t remains the same till the end of the process. Thus the number of non-adopters at the terminal date $t = N$ is

$$N_{na}^N = (1-p) \cdot \min(t^*, N).$$

Finally, the total payoff normalized to the size of the population, N, is

$$W = (N_a^N)/N = \begin{cases} p, & \text{if } k(2p-1) < 2, \\ 1 - \frac{2(1-p)}{(2p-1)k}, & \text{otherwise.} \end{cases} \tag{B.6}$$

Share of UP cascades, S_{up},($S_{down} = 0$ as have been mentioned above) and total outcome, W, as functions of k are shown at Figure 3.2 ($p = 0.75$).

Appendix C

Belief updating

C.1 Equilibria of the 1-period game

Table C.1: EP's expected payoffs, if he chooses *Agree*

VC's strategy	***Agree, E=h***		***Agree, E=l***	
	hH	***hL***	***lH***	***lL***
AAAA	$2P_{hH}-1$	$2P_{hL}-1$	$2P_{lH}-1$	$2P_{lL}-1$
AAAD	$(1+q)P_{hH}$ $-q$	$(1+q)P_{hL}$ $-q$	$(2-q)P_{lH}$ $-(1-q)$	$(2-q)P_{lL}$ $-(1-q)$
AADA	$(2-q)P_{hH}$ $-(1-q)$	$(2-q)P_{hL}$ $-(1-q)$	$(1+q)P_{lH}$ $-q$	$(1+q)P_{lL}$ $-q$
AADD	P_{hH}	P_{hL}	P_{lH}	P_{lL}
ADAA	$(1+q)P_{hH}-1$	$(1+q)P_{hL}-1$	$(2-q)P_{lH}-1$	$(2-q)P_{lL}-1$
ADAD	$q(2P_{hH}-1)$	$q(2P_{hL}-1)$	$(1-q)(2P_{lH}-1)$	$(1-q)(2P_{lL}-1)$
ADDA	$P_{hH}-(1-q)$	$P_{hL}-(1-q)$	$P_{lH}-q$	$P_{lL}-q$
ADDD	qP_{hH}	qP_{hL}	$(1-q)P_{lH}$	$(1-q)P_{lL}$
DAAA	$(2-q)P_{hH}-1$	$(2-q)P_{hL}-1$	$(1+q)P_{lH}-1$	$(1+q)P_{lL}-1$
DAAD	$P_{hH}-q$	$P_{hL}-q$	$P_{lH}-(1-q)$	$P_{lL}-(1-q)$
DADA	$(1-q)(2P_{hH}-1)$	$(1-q)(2P_{hL}-1)$	$q(2P_{lH}-1)$	$q(2P_{lL}-1)$
DADD	$(1-q)P_{hH}$	$(1-q)P_{hL}$	qP_{lH}	qP_{lL}
DDAA	$P_{hH}-1$	$P_{hL}-1$	$P_{lH}-1$	$P_{lL}-1$
DDAD	$-q(1-P_{hH})$	$-q(1-P_{hL})$	$-(1-q)(1-P_{lH})$	$-(1-q)(1-P_{lL})$
DDDA	$-(1-q)(1-P_{hH})$	$-(1-q)(1-P_{hL})$	$-q(1-P_{lH})$	$-q(1-P_{lL})$
DDDD	0	0	0	0

Table C.2: VC's expected payoffs if he chooses *Agree*

EP's strategy	***Agree, E=h***		***Agree, E=l***	
	Hh	***Hl***	***Lh***	***Ll***
AAAA	$2Q_{Hh}-1$	$2Q_{Hl}-1$	$2Q_{Lh}-1$	$2Q_{Ll}-1$
AAAD	$(1+p)Q_{Hh}$ $-p$	$(1+p)Q_{Hl}$ $-p$	$(2-p)Q_{Lh}$ $-(1-p)$	$(2-p)Q_{Ll}$ $-(1-p)$
AADA	$(2-p)Q_{Hh}$ $-(1-p)$	$(2-p)Q_{Hl}$ $-(1-p)$	$(1+p)Q_{Lh}$ $-p$	$(1+p)Q_{Ll}$ $-p$
AADD	Q_{Hh}	Q_{Hl}	Q_{Lh}	Q_{Ll}
ADAA	$(1+p)Q_{Hh}-1$	$(1+p)Q_{Hl}-1$	$(2-p)Q_{Lh}-1$	$(2-p)Q_{Ll}-1$
ADAD	$p(2Q_{Hh}-1)$	$p(2Q_{Hl}-1)$	$(1-p)(2Q_{Lh}-1)$	$(1-p)(2Q_{Ll}-1)$
ADDA	$Q_{Hh}-(1-p)$	$Q_{Hl}-(1-p)$	$Q_{Lh}-p$	$Q_{Ll}-p$
ADDD	pQ_{Hh}	pQ_{Hl}	$(1-p)Q_{Lh}$	$(1-p)Q_{Ll}$
DAAA	$(2-p)Q_{Hh}-1$	$(2-p)Q_{Hl}-1$	$(1+p)Q_{Lh}-1$	$(1+p)Q_{Ll}-1$
DAAD	$Q_{Hh}-p$	$Q_{Hl}-p$	$Q_{Lh}-(1-p)$	$Q_{Ll}-(1-p)$
DADA	$(1-p)(2Q_{Hh}-1)$	$(1-p)(2Q_{Hl}-1)$	$p(2Q_{Lh}-1)$	$p(2Q_{Ll}-1)$
DADD	$(1-p)Q_{Hh}$	$(1-p)Q_{Hl}$	pQ_{Lh}	pQ_{Ll}
DDAA	$Q_{Hh}-1$	$Q_{Hl}-1$	$Q_{Lh}-1$	$Q_{Ll}-1$
DDAD	$-p(1-Q_{Hh})$	$-p(1-Q_{Hl})$	$-(1-p)(1-Q_{Lh})$	$-(1-p)(1-Q_{Ll})$
DDDA	$-(1-p)(1-Q_{Hh})$	$-(1-p)(1-Q_{Hl})$	$-p(1-Q_{Lh})$	$-p(1-Q_{Ll})$
DDDD	0	0	0	0

Table C.3: Equilibria in 1-period game

Equilibrium	P_{hv}	P_{lv}	Q_{He}	Q_{Le}
(AAAA, AAAA)	$P_{hH} \geq P_{hL} > 1/2$	$P_{lH} \geq P_{lL} > 1/2$	$Q_{Hh} \geq Q_{Hl} > 1/2$	$Q_{Lh} \geq Q_{Ll} > 1/2$
(AAAA, AAAD)	$P_{hH} \geq P_{hL} > q/(1+q)$	$P_{lH} \geq P_{lL} > (1-q)/(2-q)$	$Q_{Hh} \geq Q_{Hl} > 1/2$	$Q_{Lh} \geq 1/2 \geq Q_{Ll}$
(AAAD, AAAA)	$P_{hH} \geq P_{hL} > 1/2$	$P_{lH} \geq 1/2 \geq P_{lL}$	$Q_{Hh} \geq Q_{Hl} > p/(1+p)$	$Q_{Lh} \geq Q_{Ll} > (1-p)/(2-p)$
(AAAA, AADD)	All	All	$Q_{Hh} \geq Q_{Hl} > 1/2$	$1/2 > Q_{Lh} \geq Q_{Ll}$
(AADD, AAAA)	$P_{hH} \geq P_{hL} > 1/2$	$1/2 > P_{lH} \geq P_{lL}$	All	All
(AAAA, ADAA)	$P_{hH} \geq P_{hL} > 1/(1+q)$	$P_{lH} \geq P_{lL} > 1/(2-q)$	$Q_{Hh} \geq 1/2 \geq Q_{Hl}$	$Q_{Lh} \geq Q_{Ll} > 1/2$
(ADAA, AAAA)	$P_{hH} \geq 1/2 \geq P_{hL}$	$P_{lH} \geq P_{lL} > 1/2$	$Q_{Hh} \geq Q_{Hl} > 1/(1+p)$	$Q_{Lh} \geq Q_{Ll} > 1/(2-p)$
(AAAA, ADAD)	$P_{hH} \geq P_{hL} > 1/2$	$P_{lH} \geq P_{lL} > 1/2$	$Q_{Hh} \geq 1/2 \geq Q_{Hl}$	$Q_{Lh} \geq 1/2 \geq Q_{Ll}$
(ADAD, AAAA)	$P_{hH} \geq 1/2 \geq P_{hL}$	$P_{lH} \geq 1/2 \geq P_{lL}$	$Q_{Hh} \geq Q_{Hl} > 1/2$	$Q_{Lh} \geq Q_{Ll} > 1/2$
(AAAA, ADDD)	All	All	$Q_{Hh} \geq 1/2 \geq Q_{Hl}$	$1/2 > Q_{Lh} \geq Q_{Ll}$
(ADDD, AAAA)	$P_{hH} \geq 1/2 \geq P_{hL}$	$1/2 > P_{lH} \geq P_{lL}$	All	All
(AAAD, AAAD)	$P_{hH} \geq P_{hL} > q/(1+q)$	$P_{lH} \geq (1-q)/(2-q) \geq P_{lL}$	$Q_{Hh} \geq Q_{Hl} > p/(1+p)$	$Q_{Lh} \geq (1-p)/(2-p) \geq Q_{Ll}$
(AAAD, ADAA)	$P_{hH} \geq P_{hL} > 1/(1+q)$	$P_{lH} \geq 1/(2-q) \geq P_{lL}$	$Q_{Hh} \geq p/(1+p) \geq Q_{Hl}$	$Q_{Lh} \geq Q_{Ll} > (1-p)/(2-p)$
(ADAA, AAAD)	$P_{hH} \geq q/(1+q) \geq P_{hL}$	$P_{lH} \geq P_{lL} > (1-q)/(2-q)$	$Q_{Hh} \geq Q_{Hl} > 1/(1+p)$	$Q_{Lh} \geq 1/(2-p) \geq Q_{Ll}$
(AAAD, ADAD)	$P_{hH} \geq P_{hL} > 1/2$	$P_{lH} \geq 1/2 \geq P_{lL}$	$Q_{Hh} \geq p/(1+p) \geq Q_{Hl}$	$Q_{Lh} \geq (1-p)/(2-p) \geq Q_{Ll}$
(ADAD, AAAD)	$P_{hH} \geq q/(1+q) \geq P_{hL}$	$P_{lH} \geq (1-q)/(2-q) \geq P_{lL}$	$Q_{Hh} \geq Q_{Hl} > 1/2$	$Q_{Lh} \geq 1/2 \geq Q_{Ll}$
(ADAA, ADAA)	$P_{hH} \geq 1/(1+q) \geq P_{hL}$	$P_{lH} \geq P_{lL} > 1/(2-q)$	$Q_{Hh} \geq 1/(1+p) \geq Q_{Hl}$	$Q_{Lh} \geq Q_{Ll} > 1/(2-p)$
(ADAA, ADAD)	$P_{hH} \geq 1/2 \geq P_{hL}$	$P_{lH} \geq P_{lL} > 1/2$	$Q_{Hh} \geq 1/(1+p) \geq Q_{Hl}$	$Q_{Lh} \geq 1/(2-p) \geq Q_{Ll}$
(ADAD, ADAA)	$P_{hH} \geq 1/(1+q) \geq P_{hL}$	$P_{lH} \geq 1/(2-q) > P_{lL}$	$Q_{Hh} \geq 1/2 \geq Q_{Hl}$	$Q_{Lh} \geq Q_{Ll} > 1/2$
(ADAD, ADAD)	$P_{hH} \geq 1/2 \geq P_{hL}$	$P_{lH} \geq 1/2 \geq P_{lL}$	$Q_{Hh} \geq 1/2 \geq Q_{Hl}$	$Q_{Lh} \geq 1/2 \geq Q_{Ll}$

C.2 Beliefs updating

Before we proceed with the discussion on how the beliefs are formed, let us redefine conditions for equilibria (Table C.3) in the terms of 'likelihood ratios'

Likelihood ratios Let P_E^t be entrepreneur's belief that $V = H$ *prior* to the private signal, given that the state of the technology $E \in \{h, l\}$. Similarly, Q_V^t will be venture capitalist's belief that $E = h$ before he receives his private signal, conditional on V. We can define "likelihood ratios" A_t, B_t, C_t, and D_t as

$$P_h^t = \frac{1}{1 + A_t}, \qquad Q_H^t = \frac{1}{1 + C_t},$$
$$P_l^t = \frac{1}{1 + B_t}, \qquad Q_L^t = \frac{1}{1 + D_t}.$$

At time $t = 0$, by assumption $P_h^0 = P_l^0 = Q_H^0 = Q_L^0 = \frac{1}{2}$, Therefore, initial values for the likelihood ratios are $A_0 = B_0 = C_0 = D_0 = 1$.

Let us denote history up to time t as I_t. According to Bayes' formula we can write P_h^t as

$$\Pr(V = H|I_t) = \frac{\Pr(I_t|V = H)\Pr(V = H)}{\Pr(I_t|V = H)\Pr(V = H) + \Pr(I_t|V = L)\Pr(V = L)} = \frac{1}{1 + \frac{\Pr(I_t|V=L)\Pr(V=L)}{\Pr(I_t|V=H)\Pr(V=H)}}$$

Given that at time $t = 0$, priors $\Pr(V = H)$ and $\Pr(V = L)$ are equal it follows that

$$P_h^t = \frac{1}{1 + \frac{\Pr(I_t|V=L)}{\Pr(I_t|V=H)}}.$$

Comparing this expression with the definition of A_t we find that

$$A_t = \frac{\Pr(I_t|V = L)}{\Pr(I_t|V = H)}.$$

in other words, A_t describes how well the history I_t can be explained with two possible alternatives $V = L$ or $V = H$.

Remark: From the definition of A_t one can also find that

$$A_t = \frac{1 - P_h^t}{P_h^t} \equiv \frac{1 - \Pr(V = H|I_t)}{\Pr(V = H|I_t)} = \frac{\Pr(V = L|I_t)}{\Pr(V = H|I_t)}.$$

Posteriors in likelihood ratios Let $E = h$. Consider an entrepreneur, his prior is P_h^t, and the corresponding likelihood ratio A_t. Suppose that he receives private signal $v = H$, i.e. entrepreneur's type is hH. According to Bayes' formula the posterior is

$$P_{hH}^t \equiv \Pr(V = H|I_t, v = H) = \frac{\Pr(v = H|I_t, V = H)\Pr(V = H|I_t)}{\Pr(v = H|I_t, V = H)\Pr(V = H|I_t) + \Pr(v = H|I_t, V = L)\Pr(V = L|I_t)} = \frac{1}{1 + \frac{\Pr(v=H|I_t,V=L)\Pr(V=L|I_t)}{\Pr(v=H|I_t,V=H)\Pr(V=H|I_t)}}.$$

Since the private signal v does not depend on the history and is determined only by the market prospects of the technology, V (see Table with signal probabilities), and taking into account the Remark above this expression can be rewritten as

$$P_{hH}^t = \frac{1}{1 + A_t \frac{1-p}{p}}$$

Similarly to the case of the hH we can write down posteriors for other types of entrepreneurs

$$P_{hH}^t = \frac{1}{1 + A_t \frac{1-p}{p}}, \qquad P_{lH}^t = \frac{1}{1 + B_t \frac{1-p}{p}},$$
$$P_{hL}^t = \frac{1}{1 + A_t \frac{p}{1-p}}, \qquad P_{lH}^t = \frac{1}{1 + B_t \frac{p}{1-p}}.$$

We can also write down posteriors for venture capitalists' beliefs

$$Q_{Hh}^t = \frac{1}{1 + C_t \frac{1-q}{q}} \qquad Q_{Lh}^t = \frac{1}{1 + D_t \frac{1-q}{q}},$$
$$Q_{Hl}^t = \frac{1}{1 + C_t \frac{q}{1-q}} \qquad Q_{Ll}^t = \frac{1}{1 + D_t \frac{q}{1-q}}.$$

Equilibria in likelihood ratios Let us consider an example. Equilibrium ($ADAD$,$AAAD$) requires

$$P_{hH} \geq \frac{q}{1+q} \geq P_{hL}, \qquad Q_{Hh} \geq Q_{Hl} > \frac{1}{2},$$
$$P_{lH} \geq \frac{1-q}{2-q} \geq P_{lL}, \qquad Q_{Ll} \geq \frac{1}{2} \geq Q_{Ll}.$$

Applying the expressions for posterior beliefs for hH and hL via likelihood ratio A_t the first of the inequalities can be rewritten as

$$\frac{1}{1+A_t\frac{1-p}{p}} \geq \frac{1}{1+\frac{1}{q}} \geq \frac{1}{1+A_t\frac{p}{1-p}}$$

or

$$A_t\frac{1-p}{p} \leq \frac{1}{q} \leq A_t\frac{p}{1-p} \Leftrightarrow \frac{1-p}{p} \leq qA_t \leq \frac{p}{1-p}.$$

In the same way we can obtain the conditions for B_t, C_t, and D_t.

Table C.4 sums up the conditions for the equilibria in 1-period game in terms of the likelyhood ratios.

Table C.4: Equilibria in likelihood ratios

Equilibrium	A_t	B_t	C_t	D_t
(AAAA, AAAA)	$A < \frac{1-p}{p}$	$B < \frac{1-p}{p}$	$C < \frac{1-q}{q}$	$D < \frac{1-q}{q}$
(AAAA, AAAD)	$qA < \frac{1-p}{p}$	$(1-q)B < \frac{1-p}{p}$	$C < \frac{1-q}{q}$	$\frac{1-q}{q} \leq D \leq \frac{q}{1-q}$
(AAAD, AAAA)	$A < \frac{1-p}{p}$	$\frac{1-p}{p} \leq B \leq \frac{p}{1-p}$	$pC < \frac{1-q}{q}$	$(1-p)D < \frac{1-q}{q}$
(AAAA, AADD)	All	All	$C < \frac{1-q}{q}$	$D > \frac{q}{1-q}$
(AADD, AAAA)	$A < \frac{1-p}{p}$	$B > \frac{p}{1-p}$	All	All
(AAAA, ADAA)	$\frac{A}{q} < \frac{1-p}{p}$	$\frac{B}{1-q} < \frac{1-p}{p}$	$\frac{1-q}{q} \leq C \leq \frac{q}{1-q}$	$D < \frac{1-q}{q}$
(ADAA, AAAA)	$\frac{1-p}{p} \leq A \leq \frac{p}{1-p}$	$B < \frac{1-p}{p}$	$\frac{C}{p} < \frac{1-q}{q}$	$\frac{D}{1-p} < \frac{1-q}{q}$
(AAAA, ADAD)	$A < \frac{1-p}{p}$	$B < \frac{1-p}{p}$	$\frac{1-q}{q} \leq C \leq \frac{q}{1-q}$	$\frac{1-q}{q} \leq D \leq \frac{q}{1-q}$
(ADAD, AAAA)	$\frac{1-p}{p} \leq A \leq \frac{p}{1-p}$	$\frac{1-p}{p} \leq B \leq \frac{p}{1-p}$	$C < \frac{1-q}{q}$	$D < \frac{1-q}{q}$
(AAAA, ADDD)	All	All	$\frac{1-q}{q} \leq C \leq \frac{q}{1-q}$	$D > \frac{q}{1-q}$
(ADDD, AAAA)	$\frac{1-p}{p} \leq A \leq \frac{p}{1-p}$	$B > \frac{p}{1-p}$	All	All
(AAAD, AAAD)	$qA < \frac{1-p}{p}$	$\frac{1-p}{p} \leq (1-q)B \leq \frac{p}{1-p}$	$pC < \frac{1-q}{q}$	$\frac{1-q}{q} \leq (1-p)D \leq \frac{q}{1-q}$
(AAAD, ADAA)	$\frac{A}{q} < \frac{1-p}{p}$	$\frac{1-p}{p} \leq \frac{B}{1-q} \leq \frac{p}{1-p}$	$\frac{1-q}{q} \leq pC \leq \frac{q}{1-q}$	$(1-p)D < \frac{1-q}{q}$
(ADAA, AAAD)	$\frac{1-p}{p} \leq qA \leq \frac{p}{1-p}$	$(1-q)B < \frac{1-p}{p}$	$\frac{C}{p} < \frac{1-q}{q}$	$\frac{1-q}{q} \leq \frac{D}{1-p} \leq \frac{q}{1-q}$
(AAAD, ADAD)	$A < \frac{1-p}{p}$	$\frac{1-p}{p} \leq B \leq \frac{p}{1-p}$	$\frac{1-q}{q} \leq pC \leq \frac{q}{1-q}$	$\frac{1-q}{q} \leq (1-p)D \leq \frac{q}{1-q}$
(ADAD, AAAD)	$\frac{1-p}{p} \leq qA \leq \frac{p}{1-p}$	$\frac{1-p}{p} \leq (1-q)B \leq \frac{p}{1-p}$	$C < \frac{1-q}{q}$	$\frac{1-q}{q} \leq D \leq \frac{q}{1-q}$
(ADAA, ADAA)	$\frac{1-p}{p} \leq \frac{A}{q} \leq \frac{p}{1-p}$	$\frac{B}{1-q} < \frac{1-p}{p}$	$\frac{1-q}{q} \leq \frac{C}{p} \leq \frac{q}{1-q}$	$\frac{D}{1-p} < \frac{1-q}{q}$
(ADAA, ADAD)	$\frac{1-p}{p} \leq A \leq \frac{p}{1-p}$	$B < \frac{1-p}{p}$	$\frac{1-q}{q} \leq \frac{C}{p} \leq \frac{q}{1-q}$	$\frac{1-q}{q} \leq \frac{D}{1-p} \leq \frac{q}{1-q}$
(ADAD, ADAA)	$\frac{1-p}{p} \leq \frac{A}{q} \leq \frac{p}{1-p}$	$\frac{1-p}{p} \leq \frac{B}{1-q} \leq \frac{p}{1-p}$	$\frac{1-q}{q} \leq C \leq \frac{q}{1-q}$	$D < \frac{1-q}{q}$
(ADAD, ADAD)	$\frac{1-p}{p} \leq A \leq \frac{p}{1-p}$	$\frac{1-p}{p} \leq B \leq \frac{p}{1-p}$	$\frac{1-q}{q} \leq C \leq \frac{q}{1-q}$	$\frac{1-q}{q} \leq D \leq \frac{q}{1-q}$

Beliefs updating

After one-period has been played the result of the game becomes public knowledge. Now we turn to how this information can be integrated into agents' beliefs.

Let the state of the technology be high, $E = h$. Suppose, that at time t the result of the negotiations is *Result* $\in\{$*Proceed, Not Proceed*$\}$. Consider an entrepreneur, who is to update his belief that $V = H$, P_h^t, given that his prior belief is

$$P_h^t \equiv \Pr(V = H|\text{history by time } t, E = h) = \frac{1}{1 + A_t},$$

According to Bayes' formula P_h^{t+1} (conditional on the *Result*), is

$$\Pr(V = H|Result) = \frac{\Pr(Result|V = H, E = h)P_h^t}{\Pr(Result|V = H, E = h)P_h^t + \Pr(Result|V = L, E = h)(1 - P_h^t)},$$

which can be rewritten (under assumption that $\Pr(Result|V = H)P_h^t \neq 0$)as

$$P_h^{t+1} = \frac{1}{1 + \frac{\Pr(Result|V=L,E=h)}{\Pr(Result|V=H,E=h)} \frac{1-P_h^t}{P_h^t}},$$

or

$$\frac{1}{1 + A_{t+1}} = \frac{1}{1 + \frac{\Pr(Result|V=L,E=h)}{\Pr(Result|V=H,E=h)} \cdot A_t}.$$

Finally, the formula for beliefs updating (in terms of the likelyhood ratios) has the form

$$A_{t+1} = A_t \cdot \frac{\Pr(Result|V = L, E = h)}{\Pr(Result|V = H, E = h)}. \tag{C.1}$$

Similarly, we can write down the rules for updating probabilities P_l^t, Q_H^t, and Q_L^t, in terms of likelyhood ratios B_t, C_t, and D_t respectively.

$$B_{t+1} = B_t \cdot \frac{\Pr(Result|V = L, E = l)}{\Pr(Result|V = H, E = l)}, \tag{C.2}$$

$$C_{t+1} = C_t \cdot \frac{\Pr(Result|V = H, E = l)}{\Pr(Result|V = H, E = h)}, \tag{C.3}$$

$$D_{t+1} = D_t \cdot \frac{\Pr(Result|V = L, E = l)}{\Pr(Result|V = L, E = h)}. \tag{C.4}$$

Example

Suppose that an entrepreneur knows that at time t equilibrium ($ADAD$, $ADAA$) has been played, and the outcome at time t is *Proceed*. How the entrepreneur should update his beliefs?

E=h, Result=Proceed The outcome is *Proceed* may have happen only if both parties had played *Agree*.

Suppose that $V = H$. Then both entrepreneur and venture capitalist at time t must have received positive signals, $v = H$ and $e = h$, respectively. Therefore,

$$\Pr(Proceed|V = H, E = h) = \Pr(v = H, e = h|E = h, V = H) = \Pr(v = H|V = H)\Pr(e = h|E = h) = p \cdot q.$$

Suppose that $V = L$. Once again, the entrepreneur must have received positive signal, $v = H$. However, in contrast with $V = H$, in this case the venture capitalist chooses *Agree* regardless to his private signal.

$$\Pr(Proceed|V = L, E = h) = \Pr(v = H, e = any|E = h, V = L) = \Pr(v = H|V = L) = 1 - p.$$

Now we are ready to apply (C.1). The updating rule is

$$A_{t+1} = A_t \cdot \frac{1-p}{p \cdot q}$$

E=l, Result=Proceed What would be different in the analysis above if $E = l$ instead of $E = h$? Note, that regardless of value of E (h or l) the entrepreneur's decision rule (at t) stays the same, he follows his signal (his strategy is $ADAD$). The value of E does not affect the venture capitalist either, since he does not know what is the true E anyway. The only thing that is going to change is the probabilities of the signals about state of the technology, e. Practically, it means that in the previous formula we should substitute q for $(1-q)$. Then the beliefs will be updated according to

$$B_{t+1} = B_t \cdot \frac{1-p}{p \cdot (1-q)}$$

Now let them play the same equilibrium ($ADAD$, $ADAA$), but suppose that the outcome at time t happend to be *Not Proceed*. How should entrepreneurs update their beliefs?

E=h, Result=Not Proceed

$$\Pr(NotProceed|V = H, E = h) = 1 - \Pr(Proceed|V = H, E = h) = 1 - p \cdot q,$$

$$\Pr(NotProceed|V = H, E = l) = 1 - \Pr(Proceed|V = H, E = l) = p,$$

$$A_{t+1} = A_t \cdot \frac{p}{1 - p \cdot q}$$

E=l, Result=Not Proceed Substituting q for $1-q$ in the formula above we get

$$B_{t+1} = B_t \cdot \frac{p}{1 - p \cdot (1-q)}$$

In the same way, using the formulas (C.1)-(C.4) we can get updating rules for entrepreneurs' and venture capitalists' beliefs for other equilibria. Those rules are listed in Table C.5.

Table C.5: Beliefs updating rule (***P:****Result=Proceed,* ***N:****Result=Not Proceed*).

Equilibrium		A_{t+1}	B_{t+1}	C_{t+1}	D_{t+1}
(AAAA, AAAA)	***P***	A_t	B_t	C_t	D_t
	N	-	-	-	-
(AAAA, AAAD)	***P***	$A_t q$	$B_t(1-q)$	C_t	$D_t \frac{1-q}{q}$
	N	∞	∞	-	$D_t \frac{q}{1-q}$
(AAAD, AAAA)	***P***	A_t	$B_t \frac{1-p}{p}$	$C_t p$	$D_t(1-p)$
	N	-	$B_t \frac{p}{1-p}$	∞	∞
(AAAA, AADD)	***P***	0	0	C_t	-
	N	∞	∞	-	D_t
(AADD, AAAA)	***P***	A_t	-	0	0
	N	-	B_t	∞	∞
(AAAA, ADAA)	***P***	$A_t \frac{1}{q}$	$B_t \frac{1}{1-q}$	$C_t \frac{1-q}{q}$	D_t
	N	0	0	$C_t \frac{q}{1-q}$	-
(ADAA, AAAA)	***P***	$A_t \frac{1-p}{p}$	B_t	$C_t \frac{1}{p}$	$D_t \frac{1}{1-p}$
	N	$A_t \frac{p}{1-p}$	-	0	0
(AAAA, ADAD)	***P***	A_t	B_t	$C_t \frac{1-q}{q}$	$D_t \frac{1-q}{q}$
	N	A_t	B_t	$C_t \frac{q}{1-q}$	$D_t \frac{q}{1-q}$
(ADAD, AAAA)	***P***	$A_t \frac{1-p}{p}$	$B_t \frac{1-p}{p}$	C_t	D_t
	N	$A_t \frac{p}{1-p}$	$B_t \frac{p}{1-p}$	C_t	D_t
(AAAA, ADDD)	***P***	0	0	$C_t \frac{1-q}{q}$	-
	N	$A_t \frac{1}{1-q}$	$B_t \frac{1}{q}$	$C_t \frac{q}{1-q}$	D_t
(ADDD, AAAA)	***P***	$A_t \frac{1-p}{p}$	-	0	0
	N	$A_t \frac{p}{1-p}$	B_t	$C_t \frac{1}{1-p}$	$D_t \frac{1}{p}$
(AAAD, AAAD)	***P***	$A_t q$	$B_t \frac{(1-p)(1-q)}{p}$	$C_t p$	$D_t \frac{(1-p)(1-q)}{q}$
	N	∞	$B_t \frac{1-(1-p)(1-q)}{1-p}$	∞	$D_t \frac{1-(1-p)(1-q)}{1-q}$
(AAAD, ADAA)	***P***	$A_t \frac{1}{q}$	$B_t \frac{1-p}{p(1-q)}$	$C_t \frac{p(1-q)}{q}$	$D_t(1-p)$
	N	0	$B_t \frac{p}{1-p(1-q)}$	$C_t \frac{1-p(1-q)}{1-q}$	∞
(ADAA, AAAD)	***P***	$A_t \frac{(1-p)q}{p}$	$B_t(1-q)$	$C_t \frac{1}{p}$	$D_t \frac{1-q}{(1-p)q}$
	N	$A_t \frac{1-(1-p)q}{1-p}$	∞	0	$D_t \frac{q}{1-(1-p)q}$
(AAAD, ADAD)	***P***	A_t	$B_t \frac{1-p}{p}$	$C_t \frac{p(1-q)}{q}$	$D_t \frac{(1-p)(1-q)}{q}$
	N	A_t	$B_t \frac{1-(1-p)(1-q)}{1-p(1-q)}$	$C_t \frac{1-p(1-q)}{1-q}$	$D_t \frac{1-(1-p)(1-q)}{1-q}$
(ADAD, AAAD)	***P***	$A_t \frac{(1-p)q}{p}$	$B_t \frac{(1-p)(1-q)}{p}$	C_t	$D_t \frac{1-q}{q}$
	N	$A_t \frac{1-(1-p)q}{1-p}$	$B_t \frac{1-(1-p)(1-q)}{1-p}$	C_t	$D_t \frac{1-(1-p)(1-q)}{1-(1-p)q}$
(ADAA, ADAA)	***P***	$A_t \frac{1-p}{pq}$	$B_t \frac{1}{1-q}$	$C_t \frac{1-q}{pq}$	$D_t \frac{1}{1-p}$
	N	$A_t \frac{p}{1-pq}$	0	$C_t \frac{q}{1-pq}$	0
(ADAA, ADAD)	***P***	$A_t \frac{1-p}{p}$	B_t	$C_t \frac{1-q}{pq}$	$D_t \frac{1-q}{(1-p)q}$
	N	$A_t \frac{1-(1-p)q}{1-pq}$	B_t	$C_t \frac{q}{1-pq}$	$D_t \frac{q}{1-(1-p)q}$
(ADAD, ADAA)	***P***	$A_t \frac{1-p}{pq}$	$B_t \frac{1-p}{p(1-q)}$	$C_t \frac{1-q}{q}$	D_t
	N	$A_t \frac{p}{1-pq}$	$B_t \frac{p}{1-p(1-q)}$	$C_t \frac{1-p(1-q)}{1-pq}$	D_t
(ADAD, ADAD)	***P***	$A_t \frac{1-p}{p}$	$B_t \frac{1-p}{p}$	$C_t \frac{1-q}{q}$	$D_t \frac{1-q}{q}$
	N	$A_t \frac{1-(1-p)q}{1-pq}$	$B_t \frac{1-(1-p)q}{1-pq}$	$C_t \frac{1-p(1-q)}{1-pq}$	$D_t \frac{1-p(1-q)}{1-pq}$

Appendix D

Local stability of the equilibria

Each Nash-equilibrium of our model gives the saturation level the new positional good will reach for a specific parameter constellation after having been introduced in the economy. To investigate their stability we rewrite the replicator dynamics (5.5) as

$$\begin{aligned} \dot{y}_1 &= y_1(1-y_1)\Delta u_1, \\ \dot{y}_2 &= y_2(1-y_2)\Delta u_2, \end{aligned} \tag{D.1}$$

where the Δu_i is defined as in equation (5.6).

D.1 Pooling (no penetration) equilibrium: $y_1 = 0$, $y_2 = 0$.

To check if this equilibrium is (locally) stable we examine the Jacobian of the replicator dynamics (D.1) at $y_1^* = 0$, $y_2^* = 0$

$$\mathcal{J}(0,0) = \begin{pmatrix} \Delta u_1(0,0) & 0 \\ 0 & \Delta u_2(0,0) \end{pmatrix}.$$

Since $\Delta u_i(0,0) < 0$, $i = 1,2$ the determinant of the Jacobian is positive, $\det \mathcal{J}(0,0) > 0$, while the trace is negative, $\operatorname{tr} \mathcal{J}(0,0) < 0$. Thus, we can conclude that for all values of q_1 and ω for which this equilibrium exists it is a stable stationary point of the system (D.1).

D.2 Separating equilibrium: $y_1 = 0$, $y_2 = 1$.

At $y_1^* = 0$, $y_2^* = 1$ the Jacobian of (D.1) has form

$$\mathcal{J}(0,1) \;=\; \begin{pmatrix} \Delta u_1(0,1) & 0 \\ 0 & -\Delta u_2(0,1) \end{pmatrix}.$$

The determinant is positive for all q_1 and ω. The sign of the trace is

$$\text{sign}(\text{tr } \mathcal{J}) = \text{sign}(\Delta u_1(0,1) - \Delta u_1(0,2)) = \\ \text{sign}(2(1-\alpha) - 4q_1 - (\Delta w_1 - \Delta w_2).$$

For the parameters (w_i, p_i) we have chosen, once the condition for this equilibrium (5.9) hold, the sign of the trace is negative. In combination with the positive determinant it implies that the equilibrium is stable for all values of parameters q_1 and ω satisfying (5.9).

D.3 Partially mixed equilibrium: $0 \leq y_1 \leq 1$, $y_2 = 1$.

The Jacobian at the point of the equilibrium is

$$\mathcal{J}(y_1^*,1) \;=\; \begin{pmatrix} 4\alpha q_1 y_1^*(1-y_1^*) & 4q_2 y_1^*(1-y_1^*) \\ 0 & -\Delta u_2(y_1^*,1) \end{pmatrix}.$$

The sign of the determinant is determined by the sign of α:

$$\text{sign}(\det \mathcal{J})) = -\text{sign}(\alpha).$$

Sign of the trace of the Jacobian is

$$\text{sign}(\text{tr } \mathcal{J}) = \text{sign}(4\alpha q_1 y_1^*(1-y_1^*) + 4q_1 y_1^* - 2\alpha q_2 + \Delta w_2 - 2q_1).$$

For q_1 and ω satisfying (5.14) the trace of the Jacobian for $\omega < 0.5$ ($\alpha < 0$) is negative, while for $\omega > 0.5$ ($\alpha > 0$) it is positive. Taking into account that tr $(\mathcal{J})^2 > 4\det(\mathcal{J})$ we can conclude that the equilibrium is a stable node of the replicator dynamics (D.1) if $\omega < 0.5$, For $\omega > 0.5$ the equilibrium is a saddle point of the replicator dynamics, and therefore it would depend on the initial conditions whether the system would move towards the equilibrium or away from it.

D.4 Equilibrium in mixed strategies: $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$.

The Jacobian at the point of the equilibrium is

$$\mathcal{J}(y_1^*, y_2^*) = \begin{pmatrix} 4\alpha q_1 y_1^*(1-y_1^*) & 4q_2 y_1^*(1-y_1^*) \\ -4q_1 y_2^*(1-y_2^*) & 4\alpha q_2 y_2^*(1-y_2^*) \end{pmatrix}.$$

The determinante and the trace of the Jacobian at (y_1^*, y_2^*) are

$$\det \mathcal{J} = 16 q_1 q_2 y_1^*(1-y_1^*) y_2^*(1-y_2^*)(1+\alpha^2) > 0,$$
$$\text{sign}(\text{tr } \mathcal{J}) = \text{sign}(4\alpha(q_1 y_1^*(1-y_1^*) + q_2 y_2^*(1-y_2^*)) = \text{sign}(\alpha).$$

Thus, equilibrium in the mixed strategies is stable for $\omega < 0.5$ ($\alpha < 0$) and unstable for $\omega < 0.5$ ($\alpha < 0$).

Appendix E

Linear feedback function

Consider N patents indexed by $i \in \{1, \dots, N\}$. Let $c_{i,t}$ denote the number of citations the patent i received at time t. The distribution of the patents at time t is $n(t) = (n_{0,t}, n_{1,t}, \dots)$, where $n_{c,t}$ is the number of patents cited c times. Further assume that the probability of a patent i to be cited is proportional to the current value of the patented invention which is related to the number of citations the patent received, c_i, as $v_{i,t} = v(c_{i,t})$. Then the *rate equation*[1] describing the dynamics of the system is

$$n_{c,t+1} - n_{c,t} = \frac{v_{c-1}}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{c-1,t} - \frac{v_c}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{c,t}, \tag{E.1}$$

where $v_c = v(c)$, and $c_{max}(t)$ is the maximum number of citations a patent may have received by time t ($c_{max}(t) = c_{max}(0) + t$). The first term in RHS (E.1) describes an increase in the number of the patents with c citations due to citing of a patent with $(c-1)$ citations. The second term in (E.1) is a loss term due to citing of a patent with c citations. Since there is no arrival of new patents (the number of patents in the cohort is fixed), for patents with no citations ($c = 0$) the first term is equal to zero, i.e.

$$n_{0,t+1} - n_{0,t} = -\frac{v_0}{\sum_{j=0}^{c_{max}(t)} v_j n_{j,t}} n_{0,t}. \tag{E.2}$$

At time $t = 0$ the distribution of patents is

$$n(0) = (n_0^0, \dots, n_{c_{max}(0)}^0, 0, \dots).$$

Equations (E.1), (E.2), and the initial condition define the evolution of the system.

Linear feedback In case of the linear value (preferential attachment) function (6.2) we can find a closed-form solution. The sum in the denominator of the equation

[1] deterministic equation that describes evolution of expected values

(E.1) is

$$\sum_{j=0}^{c_{max}(t)} (v_0 + j) n_{j,t} = V_0 + t, \tag{E.3}$$

where V_0 is the sum of the values of the patented inventions at $t = 0$, i.e. $V_0 = v_{1,0} + \dots v_{N,0}$. Note that

$$V_0 = \sum_{j=0}^{c_{max}(0)} (v_0 + j) = v_0 N + t_0,$$

where t_0 is the total number of citations at $t = 0$, i.e. citations within the cohort.

For the sake of simplicity assume that at $t = 0$ patents have no citations ($n_0^0 = N$, and $n_c^0 = 0\ \forall c \neq 0$, i.e. $t_0 = 0$). The evolution of the system can be analyzed drawing a binomial tree: each node ot the tree represents $n_{c,t}$, the probability of transition from node (c, t) to the node $(c+1, t+1)$ is equal to the probability of citing a patent with c citations at time t. Given (E.1) it is clear that the probability to arrive to node (c, t) from the origin $(0, 0)$ does not depend on the path chosen. Taking into account this fact the further derivation of the distribution $n(t)$ becomes trivial, for $c \geq 1$ the result is

$$n_{c,t} = N \binom{t}{c} \frac{\prod_{i=0}^{c-1}(v_0 + i) \prod_{i=0}^{t-c-1}(V_0 - v_0 + i)}{\prod_{i=0}^{t-1}(V_0 + i)} = \\ = N \binom{t}{c} \frac{B(V_0 + t, v_0 + c)}{B(V_0, v_0)}, \tag{E.4}$$

where $B(x, y)$ is a Legendre beta-function. For patents with no citations ($c = 0$)

$$n_{0,t} = N \prod_{i=0}^{t-1} \left(1 - \frac{v_0}{V_0 + i}\right).$$

Provided that $V_0 = v_0 N >> v_0$ we can approximate it as

$$\ln n_{0,t} = \ln N + \sum_{i=0}^{t-1} \ln \left(1 - \frac{v_0}{V_0 + i}\right) \approx \ln N - \sum_{i=0}^{t-1} \ln \frac{v_0}{V_0 + i} \approx \\ \approx \ln N - v_0 \ln \left(1 + \frac{t}{V_0}\right),$$

therefore

$$n_{0,t} = N \left(1 + \frac{t}{V_0}\right)^{-v_0}. \tag{E.5}$$

Since the linear model has only one parameter (v_0), fitting of this parameter can be done with the equation (E.5) using only data on $n_{0,t}$.

The equation (E.4) can be rewritten in a more convenient recursive form as following ($c \geq 1$)

$$\ln n_{c+1,t} = \ln n_{c,t} + \ln \frac{(t-c)(v_0+c)}{(c+1)(t-c+V_0-v_0-1)}. \tag{E.6}$$

The system of (E.5) and (E.6) provides us with the distribution of patent citations at any t.

The linearity of the process allows us to extend the solution to the general case of initial conditions. The resulting distribution is simply a superposition of the distributions, which one can derive from the analysis of binomial trees with origin in $(c, 0)$ where $0 \leq c \leq c_{max}(0)$.

Bibliography

Aghion, P., and P. Howitt (1992): "A Model of Growth through Creative Destruction," *Econometrica*, 60(2), 323–51.

Aisner, R., and J. Terkel (1992): "Ontogeny of the pine opening behavior in the Black Rat (*Rattus rattus*)," *Animal Behaviour*, 44, 327–336.

Akerlof, G. (1976): "The Economics of Caste and of the Rat Race and Other Woeful Tales," *The Quarterly Journal of Economics*, 90(4), 599–617.

Akerlof, G. A. (1980): "A Theory of Social Custom, of Which Unemployment May be One Consequence," *The Quarterly Journal of Economics*, 94(4), 749–775.

——— (1997): "Social Distance and Social Decisions," *Econometrica*, 65(5), 1005–1027.

Albert, R., and A. Barabasi (2002): "Statistical mechanics of complex networks," *Reviews of Modern Physics*, 74, 47–97.

Anderson, L. R., and C. A. Holt (1997): "Information Cascades in the Laboratory," *The American Economic Review*, 87(5), 847–862.

Arthur, W. B. (1989): "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *Economic Journal*, 99(394), 116–31.

——— (1994): *Increasing returns and path dependence in the economy*. University of Michigan Press, Ann Arbor.

Arthur, W. B., Y. M. Ermoliev, and Y. M. Kaniovski (1983): "A Generalized Urn Problem and its Applications," *Kibernetika*, 19, 49–56.

Avery, C., and P. Zemsky (1998): "Multidimensional Uncertainty and Herd Behavior in Financial Markets," *American Economic Review*, 88(4), 724–748.

Bagwell, L. S., and B. D. Bernheim (1996): "Veblen Effects in a Theory of Conspicuous Consumption," *The American Economic Review*, 86(3), 349–373.

Bala, V., and S. Goyal (1994): "The Birth of a New Market," *The Economic Journal*, 104(423), 282–290.

Bala, V., and S. Goyal (1998): "Learning from Neighbours," *Review of Economic Studies*, 65(3), 595–621.

——— (2001): "Conformism and Diversity under Social Learning," *Economic Theory*, 17(1), 101–20.

Balconi, M., S. Breschi, and F. Lissoni (2004): "Networks of Inventors and the Role of Academia: An Exploration of Italian Patent Data," *Research Policy*, 33(1), 127–45.

Bandura, A. (1976): "Social learning theory," in *Behavioral approaches to therapy.*, ed. by J. T. Spence, R. C. Carson, and J. W. Thibaut. General Learning Press, Morristown, NJ.

——— (1977): *Social learning theory.* Prentice-Hall, Englewood Cliffs, NJ.

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Barabasi, A. (2002): *Linked. The New Science of Networks.* Perseus, New York.

Bassanini, A. P., and G. Dosi (1999): "Competing Technologies, Technological Monopolies, and the Rate of Convergence to a Stable Market Structures," *mimeo.*

Baum, J. A. C., S. X. Li, and J. M. Usher (2000): "Making the Next Move: How Experiential and Vicarious Learning Shape the Locations of Chains' Acquisitions," *Administrative Science Quarterly*, 45(4), 766–801.

Becker, G. S. (1976): "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology," *Journal of Economic Literature*, 14(3), 817–826.

——— (1991): "A Note on Restaurant Pricing and Other Examples of Social Influences on Price," *Journal of Political Economy*, 99(5), 1109–1116.

Bernardo, A. E., and I. Welch (2001): "Novelty, Preferences, and Fashion: When Goods Are Unsettling," *Journal of Economics & Management Strategy*, 10(3), 301330.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100(5), 992–1026.

——— (1998): "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives*, 12(3), 151–170.

Birchenhall, C. (1995): "Modular Technical Change and Genetic Algorithms," *Computational Economics*, 8(3), 233–53.

Blackmore, S. J. (1998): "Imitation and the Definition of a Meme," *Journal of Memetics - Evolutionary Models of Information Transmission*, 2(2), 159–170.

——— (2000): "The power of memes," *Scientific American*, 283(4), 52–61.

Bolton, P., and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67(2), 349–374.

Bourdieu, P. (1984): *Distinction*. Harvard University Press., Cambridge, MA.

Branscomb, L. M., and P. E. Auerswald (2001): *Taking technical risks: How innovators, executives, and investors manage high-tech risks*. MIT Press, Cambridge and London.

Breschi, S., and F. Lissoni (2004): "Knowledge networks from patent data: Methodological issues and research targets," *CESPRI Working Papers*, 150.

Burns, L. R., and D. R. Wholey (1993): "Adoption and Abandonment of Matrix Management Programs: Effects of Organizational Characteristics and Interorganizational Networks," *Academy of Management Journal*, 36(1), 106–138.

Busenitz, L., and J. Barney (1997): "Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making," 12, 9–30.

Campbell, C. (1995): "The Sociology of Consumption," in *Acknowledging Consumption*, ed. by D. Miller, pp. 96–126. Routledge, London.

Caplin, A., and J. Leahy (1994): "Business as Usual, Market Crashes, and Wisdom After the Fact," *The American Economic Review*, 84(3), 548–565.

Chung, F., S. Handjani, and D. Jungreis (2003): "Generalizations of Polyas urn problem," *Annals of Combinatorics*, 7, 141–154.

Cipolla, C. (1967): *Clocks and Culture*. Norton, London.

Conell, C., and S. Cohn (1995): "Learning from Other People's Actions: Environmental Variation and Diffusion in French Coal Mining Strikes, 1890-1935," *American Journal of Sociology*, 101(2), 366–403.

Congleton, R. D. (1989): "Efficient Status Seeking: Externalities, and the Evolution of Status Games," *Journal of Economic Behavior and Organization*, 11(2), 175–90.

Cooper, B. (2000): "Modelling Research and Development: How Do Firms Solve Design Problems?," *Journal of Evolutionary Economics*, 10(4), 395–413.

Corneo, G., and O. Jeanne (1997a): "Conspicuous Consumption, Snobbism and Conformism," *Journal of Public Economics*, 66(1), 55–71.

——— (1997b): "Snobs, Bandwagons, and the Origin of Social Customs in Consumer Behavior," *Journal of Economic Behavior and Organization*, 32(3), 333–47.

Cowan, R. (1990): "Nuclear Power Reactors: A Study in Technological Lock-in," *Journal of Economic History*, 50(3), 541–567.

Cowan, R., W. Cowan, and P. Swann (1997): "A Model of Demand with Interactions among Consumers," *International Journal of Industrial Organization*, 15(6), 711–32.

Cowan, R., and D. Foray (2002): "Evolutionary Economics and the Counterfactual Threat: On the Nature and Role of the Counterfactual History as an Empirical Tool in Economics," *Journal of Evolutionary Economics*, 12(5), 539–62.

Cressman, R. (1995): "Evolutionary Game Theory with Two Groups of Individuals," *Games and Economic Behavior*, 11(2), 237–53.

Cyert, R. M., and J. G. March (1963): *A behavioral theory of the firm*, Prentice Hall international series in management. Prentice Hall, Englewood Cliffs.

David, P. A. (1985): "Clio and the Economics of QWERTY," *The American Economic Review*, 75(2, Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association), 332–337.

Davies, S. (1979): *The Diffusion of Process Innovations*. Cambridge University Press, Cambridge.

Davis, T. R. V., and F. Luthans (1980): "A Social Learning Approach to Organizational Behavior," *Academy of Management Review*, 5(2), 281–290.

Dawid, H. (1999): *Adaptive learning by genetic algorithms: Analytical results and applications to economic models*. Springer, Heidelberg and New York:, second edition edn.

Dawkins, R. (1976): *The Selfish Gene*. Oxford University Press, Oxford.

de Geus, A. (1997): *The living company*. Harvard Business School Press, Boston.

de Vries, J. (1993): "Between Purchasing Power and the World of Goods: Understanding the Household Economy in Early Modern Europe," in *Consumption and the World of Goods*, ed. by J. Brewer, and R. Porter, pp. 85–132. Routledge, London and New York.

——— (1994): "The Industrial Revolution and the Industrious Revolution," *Journal of Economic History*, 54(2), 249–270.

DiMaggio, P. J., and W. W. Powell (1983): "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields," *American Sociological Review*, 48(2), 147–160.

Dosi, G. (1982): "Technical paradigms and technological trajectories: A suggested interpretation of the determinants of technical change," *Research Policy*, II(3), 147162.

Duesenberry, J. S. (1949): *Income, Saving and the Theory of Consumer Behavior*. Cambridge University Press, London.

Ellison, G., and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101(4), 612–643.

——— (1995): "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*, 110(1), 93–125.

Elster, J. (1989): "Social Norms and Economic Theory," *The Journal of Economic Perspectives*, 3(4), 99–117.

Farrell, J., and G. Saloner (1986): "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *The American Economic Review*, 76(5), 940–955.

Fehr, E., and S. Gächter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14(3), 159–81.

Fisher, J., and R. Hinde (1949): "The opening of milk bottles by birds," *British Birds*, 42, 347–357.

Fisher, J. C., and R. Pry (1971): "A Simple Substitution Model of Technological Change," *Technological Forecasting and Social Change*, 3, 75–88.

Fligstein, N. (1985): "The Spread of the Multidivisional Form Among Large Firms, 1919-1979," *American Sociological Review*, 50(3), 377–391.

Frank, R. H. (1985): "The Demand for Unobservable and Other Nonpositional Goods," *The American Economic Review*, 75(1), 101–116.

Frenken, K. (2004): "History, state and prospects of evolutionary models of technical change: a review with special emphasis on complexity theory," in *Complexity Focus*, ed. by J. Casti. Exystence, Turin.

Frenken, K., and A. Nuvolari (2004): "The early development of the steam engine: An evolutionary interpretation using complexity theory," *Industrial and Corporate Change*, 13(2), 419–450.

Geroski, P. A. (2000): "Models of Technology Diffusion," *Research Policy*, 29(4-5), 603–25.

Gilboy, E. W. (1932): "Demand as a Factor in the Industrial Revolution," in *The Causes of the Industrial Revolution in England*, ed. by A. Cole. Harvard University Press, Cambridge, MA.

Gompers, P. A., and J. Lerner (1999): *The venture capital cycle*. MIT Press, Cambridge and London.

Graff, G. D. (2003): "Observing Technological Trajectories in Patent Data: Empirical Methods to Study the Emergence and Growth of New Technologies," *American Journal of Agricultural Economics*, 85(5), 1266–74.

Griliches, Z. (1957): "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25(4), 501–522.

——— (1981): "Market value, R&D, and patents," *Economics Letters*, 7(2), 183–187.

——— (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28(4), 1661–1707.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," *NBER Working Papers*, 8498.

——— (2005a): "Market Value and Patent Citations: A First Look," *Rand Journal of Economics*, 36, 16–38.

——— (2005b): "Market Value and Patent Citations: A First Look," *Rand Journal of Economics*, 36, 16–38.

Hall, B. H., and M. Trajtenberg (2004): "Uncovering GPTs with Patent Data," *NBER Working Papers*, 10901.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999): "Citation Frequency and the Value of Patented Inventions," *Review of Economics and Statistics*, 81(3), 511–15.

Harhoff, D., F. M. Scherer, and K. Vopel (1997): "Exploring the tail of patent value distributions," *Center for European Economic Research Discussion Paper*, 97-30.

——— (2003): "Citations, Family Size, Opposition and the Value of Patent Rights," *Research Policy*, 32(8), 1343–63.

Hartmann, G., and M. Myers (2001): "Technical Risk, Product Specifications, and Market Risk," in *Taking technical risks: How innovators, executives, and investors manage high-tech risks*, ed. by L. M. Branscomb, and P. E. Auerswald. MIT Press, Cambridge and London.

Hartwell, R. (1965): "The Causes of the Industrial Revolution: An Essay in Methodology," *The Economic History Review*, 18(1), 164–182.

Haunschild, P. R., and A. S. Miner (1997): "Modes of Interorganizational Imitation: The Effects of Outcome Salience and Uncertainty," *Administrative Science Quarterly*, 42(3), 472–500.

Haveman, H. A. (1993): "Follow the Leader: Mimetic Isomorphism and Entry Into New Markets," *Administrative Science Quarterly*, 38(4), 593–627.

Hayakawa, H., and Y. Venieris (1977): "Consumer Interdependence via Reference Groups," *The Journal of Political Economy*, 85(3), 599–616.

Henderson, R., A. B. Jaffe, and M. Trajtenberg (1998): "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988," *Review of Economics and Statistics*, 80(1), 119–27.

Heyes, C. (1993): "Imitation, culture and cognition," *Animal Behaviour*, 46, 999–1010.

Heyes, C., and B. Galef Jr. (1996): *Social Lerning in Animals. The Roots of Culture*. Academic Press, San Diego.

Hill, B. (1975): "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 3(5), 1163–1173.

Houthakker, H. (1955): "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, 23, 2731.

Huang, Z., H. Chen, A. Yip, G. Ng, F. Guo, Z.-K. Chen, and M. C. Roco (2003): "Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field," *Journal of Nanoparticle Research*, 5, 333–363.

Iannaccone, L. R. (1989): "Bandwagons and the Threat of Chaos: Interpersonal Effects Revisited," *Journal of Economic Behavior and Organization*, 11(3), 431–42.

Ingold, T. (1990): "An Anthropologist Looks at Biology," *Man*, 25(2), 208–229.

Jaffe, A. B., and M. Trajtenberg (2002): *Patents, citations, and innovations: A window on the knowledge economy*. MIT Press.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2000): "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review*, 90(2), 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108(3), 577–98.

Janssen, M. A., and W. Jager (2001): "Fashions, Habits and Changing Preferences: Simulation of Psychological Factors Affecting Market Dynamics," *Journal of Economic Psychology*, 22(6), 745–72.

Jones, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *Quarterly Journal of Economics*, 120(2), 517–47.

Kauffman, S. A. (1993): *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford and New York.

Kirman, A. (1993): "Ants, Rationality, and Recruitment," *The Quarterly Journal of Economics*, 108(1), 137–156.

Kortum, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65(6), 1389–1419.

Krapivsky, P. L., S. Redner, and F. Leyvraz (2000): "Connectivity of Growing Random Networks," *Phys. Rev. Lett.*, 85(21), 4629–4632.

Lanjouw, J. O., A. Pakes, and J. Putnam (1998): "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data," *Journal of Industrial Economics*, 46(4), 405–32.

Leibenstein, H. (1950): "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," *The Quarterly Journal of Economics*, 64(2), 183–207.

MacKay, C. (1841): *Extraordinary Popular Delusions and the Madness of Crowds*, Wordsworth Reference Series. Wordsworth Editions Ltd, London. 1995.

Mansfield, E. (1961): "Technical Change and the Rate of Imitation," *Econometrica*, 29(4), 741–766.

McGroddy, J. (2001): "Raising Mice in the Elephants' Cage," in *Taking technical risks: How innovators, executives, and investors manage high-tech risks*, ed. by L. M. Branscomb, and P. E. Auerswald. MIT Press, Cambridge and London.

Meltzoff, A. (1988): "The Human Infant as Homo Imitans," in *Social learning: Psychological and Biological Perspectives*, ed. by T. Zentall, and B. Galef Jr.

Mokyr, J. (1977): "Demand vs. Supply in the Industrial Revolution," *The Journal of Economic History*, 37(4), 981–1008.

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

——— (1982): *An Evolutionary Theory of Economic Change*. Belknap Press, Cambridge, Mass. and London.

Newman, M. (2005): "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, (forthcoming).

Pakes, A. (1985): "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93(2), 390–409.

Pakes, A., and M. Schankerman (1984): "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in *R&D, Patents, and Productivity*, ed. by Z. Griliches, NBER Conference Report, pp. 73–88. University of Chicago Press, Chicago and London.

Pakes, A. S. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54(4), 755–84.

Palmer, D. A., P. D. Jennings, and X. Zhou (1993): "Late Adoption of the Multidivisional Form by Large U.S. Corporations: Institutional, Political, and Economic Accounts," *Administrative Science Quarterly*, 38(1), 100–131.

Pesendorfer, W. (1995): "Design Innovation and Fashion Cycles," *The American Economic Review*, 85(4), 771–792.

Petersen, M. A., and R. G. Rajan (2002): "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, 57(6), 2533–2570.

Pier, H., K. Rost, T. Teichert, and I. von Wartburg (2003): "H04Q-007 - Deconstructing the Technological Blob of Mobile Telecommunication. An Application of Social Network Analysis for Categorizing Technological Fields," in *ZEW Workshop "Empirical Innovation and Patenting"*, Mannheim, Germany. March 14-15.

Posner, R. A. (1997): "Social Norms and the Law: An Economic Approach," *The American Economic Review*, 87(2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association), 365–369.

Rae, J. (1905): *The Sociological Theory of Capital*. Macmillan, London.

Rao, H., H. R. Greve, and G. F. Davis (2001): "Fool's Gold: Social Proof in the Initiation and Abandonment of Coverage by Wall Street Analysts," *Administrative Science Quarterly*, 46(3), 502–526.

Rauscher, M. (1993): "Demand for Social Status and the Dynamics of Consumer Behavior," *Journal of Socio-Economics*, 22(2), 105–13.

Reinstaller, A., and B. Sanditov (2005): "Social structure and consumption: on the diffusion of consumer good innovation," *Journal of Evolutionary Economics*, 15(5), 505–31.

Rosenberg, N. (1968): "Adam Smith, Consumer Tastes, and Economic Growth," *The Journal of Political Economy*, 76(3), 361–374.

——— (1969): "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices," *Economic Development and Cultural Change*, 18(1), 1–24.

——— (1974): "Science, Invention and Economic Growth," *Economic Journal*, 84(333), 90–108.

——— (1982): "Exploring the black box: Technology, economics, and history," .

Sahal, D. (1981): *Patterns of Technological Innovation.* Addison-Wesley, New York.

Schankerman, M., and A. Pakes (1986): "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period," *Economic Journal*, 96(384), 1052–76.

Scharfstein, D. S., and J. C. Stein (1990): "Herd Behavior and Investment," *The American Economic Review*, 80(3), 465–479.

Scherer, F. M., and D. Harhoff (2000): "Technology Policy for a World of Skew-Distributed Outcomes," *Research Policy*, 29(4-5), 559–66.

Sen, A. (1985): *Commodities and Capabilities.* Oxford University Press, New Dehli.

——— (1997): "Maximization and the Act of Choice," *Econometrica*, 65(4), 745–779.

Shiller, R. J. (1989): *Market volatility.* MIT Press, Cambridge, Mass. and London.

Silverberg, G., and B. Verspagen (2004): "The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance," *MERIT/Infonomics research memoranda*, 2004-021, *forthcoming in the Journal of Econometrics.*

——— (2005): "A percolation model of innovation in complex technology spaces," *Journal of Economic Dynamics and Control*, 29(1-2), 225–244.

Simmel, G. (1957): "Fashion," *American Journal of Sociology*, 62(6), 541–58.

Smith, J. M. (1964): "Group selection and kin selection," *Nature*, 201, 114547.

Strang, D., and M. C. Still (2004): "In search of the elite: revising a model of adaptive emulation with evidence from benchmarking team," *Industrial and Corporate Change*, 13(2), 309–333.

Swann, G. M. P. (2001): "The Demand for Distinction and the Evolution of the Prestige Car," *Journal of Evolutionary Economics*, 11(1), 59–75.

Taylor, P. D. (1979): "Evolutionary Stable Strategies with two Types of Player," *Journal of Applied Probability*, 16, 76–83.

Terkel, J. (1996): "Cultural Transmission of Feeding Behavior in the Black Rat (*Rattus rattus*)," in *Social Lerning in Animals. The Roots of Culture*, ed. by C. Heyes, and B. Galef Jr. Academic Press, San Diego.

Trajtenberg, M. (1990): "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21(1), 172–87.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997): "University versus Corporate Patents: A Window on the Basicness of Invention," *Economics of Innovation and New Technology*, 5(1), 19–50.

Veblen, T. (1921): *The theory of the leisure class*, Reprints of Economic Classics. The Writings of Thorstein Veblen. AM Kelly, Fairfield, NJ. 1991.

Voigtländer, N., and H.-J. Voth (2005): "Why England? Demand, growth and inequality during the Industrial Revolution," *Universitat Pompeu Fabra, Mimeo.*

Watts, D. J. (1999): *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton University Press, Princeton.

Weibull, J. W. (1995): *Evolutionary game theory.* MIT Press, Cambridge and London.

Weisbuch, G., A. Kirman, and D. Herreiner (2000): "Market Organisation and Trading Relationships," *Economic Journal*, 110(463), 411–436.

Welch, I. (1992): "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 47(2), 695–732.

Yule, G. (1925): "A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis.," *Philos. Trans. R. Soc. London B*, 213, 2187.

Zacharakis, A., and D. Shepherd (2001): ""The Nature of Information and Overconfidence on Venture Capitalists' Decision Making,"," *Journal of Business Venturing, 2001, 16:311-332.*, 16, 311–32.