The 'problem of problem choice': A model of sequential knowledge production within scientific communities

Nicolas Carayol[θ*], **Jean-Michel Dalle**[β]

[θ] BETA, UMR CNRS 7522, Université Louis Pasteur,

61, avenue de la Forêt Noire, F-67085 Strasbourg Cedex

[β] Université Paris 6 and IMRI - Université Paris Dauphine

[*] Corresponding author: Tel: +33(0)390242104 - Fax: +33(0)390242071

<carayol@univ-tlse1.fr>

Abstract

In this paper we present an original model of sequential problem choice within scientific communities. Disciplinary knowledge is accumulated by solving problems emerging in a growing tree-like web of research areas. Knowledge production is sequential since the problems solved generate new problems that may be handled. The model allows us to study how the reward system in science influences the scientific community in stochastically selecting at each period its research agendas, and the long term resulting disciplines. We present some evidence on a decrease in the generation of new areas, a path dependency in specialization, and circumstances under which collapsing dynamics arise.

Key words: Sequential Problem Choice; Stochastic Process; Tree; Graph Theory; Scientific Knowledge; Academics; Reward System

JEL classification: A12; C63; H40; O30

Acknowledgements

The authors would like to thank R.K. Merton for having let them read an unpublished manuscript written with R.C. Merton, for his comments on an earlier version of the paper, and for having helped us track the sources of the "problem of problem choice in science". P.A. David should also be strongly acknowledged for his helpful comments at different stages of our work. The technical and non technical comments of A. Kirman were also very useful. A preliminary version of this paper was presented at the WEHIA Conference (5th Workshop on Economics with Heterogeneous Interacting Agents, GREQAM, Marseille, June 2000 15-17th), at the CREREG (University of Rennes 1) seminar , and at the LIRHE (University of Toulouse 1) seminar, under the title: "Science wells: The 'problem of problem choice' within scientific communities". A connected version was presented at the NPR-net conference ('Rethinking science Policy: Analytical Frameworks for evidenced based policy', SPRU, Brighton, March 2002), and at the EGOS conference (European Group for Organizational Studies conference, July 2001 5-7th, Lyon), under the title "Science-industry relations and Scientists' problem of problem choice". We would like to thank the participants of these seminars and conference sessions for useful comments and remarks. Remaining errors are ours.

1 Introduction

Nelson (1959) and Arrow (1962) were the first to highlight that the specific characteristics of knowledge considered as a public good result in a default in knowledge creation incentives. Consequently, the level of private investment in knowledge creation is below its optimal level. This very well known result appeared as a theoretical justification for public support of research which may (non exclusively) be undertaken by funding a specific social institution, namely the academia. In that respect, modern countries obviously support a network of public laboratories and academic researchers. Even if some econometric works have indeed shown that the social returns from public research could be quite high[1], economists naturally have begun to wonder about the microeconomic properties of the academic institution. To do so they relied on the initial contributions of the Sociology of Science. According to the early work of its founder R.K. Merton, the functioning of the academic institution, that he labels *Open Science,* relies on four "institutional imperatives", namely "universalism, communism, disinterestedness, and organized skepticism" (Merton, 1942). As argued later, these norms generate a set of effective rules which stress a specific *reward system,* in which priority is essential *(*Merton, 1957). The incentive mechanism at play may be sketched as follows. Peers collectively establish the validity and novelty of knowledge produced (peer review). The attribution of rewards is based on recognition by peers of the "moral property" on the piece of knowledge produced which increases the producer's reputation within the community ("credit"). Using these notions, Dasgupta and David (1994) have recently synthesized in an economic fashion the mechanisms at play within academia. They highlighted that its functioning has two fundamental and original economic properties. First of all, it avoids some of the asymmetric-informational problems that might otherwise arise between funding agencies and scientists in public procurement of advanced knowledge: scientists themselves are certainly the most able to carry out verification and evaluation operations in the peer-review like procedures. Secondly, since it is precisely the very action of disclosing knowledge which induces the reward (reputation or credit increase), the reward system thus creates simultaneous incentives both for knowledge creation and for its early disclosure and broad dissemination within the community. That is why this mode of

[1] For a recent and complete survey see Salter and Martin (2001).

knowledge production has been said to have very interesting efficiency properties (Arrow, 1987) and even to constitute a "first best solution" for the appropriability problem (Dasgupta and David, 1994) as it solves the dilemma between knowledge creation incentives and knowledge disclosure incentives (Stephan 1996)[2].

Nevertheless, the mechanisms described above are only part of the story. As a matter of fact, a very consubstantial feature of the Open Science organization, is the relative freedom scientists have for defining and selecting their own *research agendas*. That is the reason why one may say that Open Science is indeed a *decentralized system of knowledge production*. Thus, one may wonder how the Open Science reward system influences scientists' problem selection i.e. selection of their research agendas. And that issue is far from being marginal from scientists' points of views: not all problems are the same in their eyes. Though competition between scientists is clearly important, associated with "winner-takes-all" rules and "waiting and racing games" issues (Dasgupta and David, 1987; Reinganum, 1989), it is only second, while the first and most important decision a scientist has to take is the choice of which research area he or she will investigate, of his own research agenda (Dasgupta and Maskin, 1987). How the reward system affects problem selection is also crucial for society because it certainly shapes the pace of scientific knowledge production. In wondering about the economic properties of the *research agenda determination process* within Open Science we are meeting the issue which is usually referred to in the sociology of science as the "problem of problem choice" (Merton, 1957; Zuckermann, 1978; Ziman, 1987)[3]. The only 'economic-like' contribution to that field is the one of Merton and Merton (1989) who built an optimal control model in which researchers' efforts are dedicated to solving a given set of problems. This model aims to compare the decentralized allocation of research efforts with the optimal one[4]. Nevertheless, their model is limited to

[2] Of course, many problems still arise and it is not possible to derive from this statement that the decentralized allocation of research efforts induced by the specific reward system of science is *per se* optimal. This observation leaves the room for an *Imperfect Economics of Science* to come (for a first investigation see Carayol, 2001).

[3] The history of this issue began with Peirce (1896) who provided the first formal model of the optimal allocation of research between projects characterized by different levels of utility and risk. Merton (1957) first noticed the importance of competition and the reward system on problem choice. Polanyi (1962) suggested that science is a self-organized and indeed efficient institution for orienting the scientists' attention and research efforts in a decentralized manner. He suggested that scientists are guided by an "invisible hand for ideas" - that science is a well designed social institution because it allows for an optimal allocation of research efforts.

[4] Unfortunately, this paper which was initially supposed to appear (and announced) in *Rationality and Society* in 1989 has been withdrawn from publication at the time. The available version is still uncomplete and most

situations in which the set of problems to be solved is fixed, the problems are independent[5], and the rewards for solving each of them are exogenously fixed and *ex ante* specified.

In this paper we present a dedicated model of the sequential determination of research agendas and subsequent dynamic knowledge production within open science. The model is designed to capture two features of the process of collective knowledge creation which are essential for analyzing scientific knowledge production within academic communities. They are the following:

i) the set of problems to be solved is not fixed: rather our aim is to account for the idea that new problems are generated by previously created knowledge, that is by problems solved in the past. This observation highlights the sequential nature of knowledge production;

ii) the rewards for solving problems are not exogenously specified: the allocation of research efforts on a set of handleable problems at any given period of time is derived from both the generic *motivations* of scientists and how they presently anticipate the recompenses for solving each problem defined according to the specific reward system in science.

In order to take these two features into account we propose a model using graph theoretical principles: we assume that *disciplines* (a term by which we simply mean the *accumulated knowledge of a scientific discipline*) are designed as a *web of research areas* (i.e. a graph, the nodes of which are designating research areas)[6]. The research areas are the unitary level of scientific knowledge organization. Each one is simply defined by its *location* in the graph and its *level of improvement*, that is the *number of problems* that have been solved within that area (which can also be understood as the past accumulation of knowledge there). For the sake of simplicity, this web is assumed to be a *tree*-like graph. Thus we retain the classical, even if somewhat misleading, representation of scientific knowledge as a *tree of knowledge*[7]. Even if this simplification is far from being perfect, it makes it possible to clearly organize areas

results are still unavailable.

[5] There is no externality: the solution of one problem does not influence the solving of another problem.

[6] The initial idea of scientific knowledge as a web of "theories" is to be found in Kuhn (1962). Even if very different in its conception, see also Weitzman (1998) for one of the first contributions 'discretizing' knowledge.

[7] See Machlup (1982) for a history of this notion which dates back to Lull, Bacon and Comte. Some first elements can also be found in Cournot (1861) where the idea of a *branching* evolution of knowledge in the sciences and industry is introduced. See also Ziman (1984) for a rationale about its applicability to the usual scientific classification systems.

according to their level of specialization, through their geodesic distance to the root which is assumed to be the locus of the most general knowledge. The set of already existing research areas (nodes) and possible ones, their structure and their respective levels of improvement, altogether constitute the present state of the discipline. According to the sequential character of knowledge production highlighted in *i*), the state of the system is assumed to directly give the set of attainable problems at each period: solving the next problem to hand on the existing research areas or exploring new research areas. The latter is modeled by introducing the opportunity that any already improved research area leads to the potential creation of a new research area (which one could view as a new *leaf* of the tree) by solving its first problem.

Thus the remaining issue, stated in *ii*), becomes - in the context of our model - how agents strategically allocate their *research efforts* over the set of handleable problems. Their choices are obviously mainly a function of the net expected reward for solving each. In science, the rewarding process is reputation based: scientists seek after "credit". It is known that such symbolic rewards are positively correlated with monetary ones (Stephan, 1996). To get such rewards scientists should attract other scientists attention by producing what Cohen et al. (1998) call "foundational knowledge". Thanks to the new computing techniques and availability of large scientific database, citations are now often used as a common weighting parameter of the importance of scientific contributions[8]. In addition to publications, scientists' CVs are increasingly mentioning the number of citations they received[9]. All this suggests that scientists clearly try to anticipate such expected returns of their marginal research efforts while choosing their research agendas. It opens the door to considering scientists, while choosing their research agendas, as '*credit seekers'*. In order to capture that, we introduce a specific *reward function* determining how agents associate at a given moment in time an expected reward to each attainable problem. Incentives to performing research on research areas are assumed to be a function of three variables that have been highlighted by the empirical literature on problem choice in science[10]. The first variable indicates how much the research associated with a given problem

[8]However, citation counts should be taken carefully especially if one wishes to compare the quality of papers appearing in different fields.

[9]This kind of reward is crucial, because it is encouraging scientists to spend more time and support higher risks that important but difficult problems induce, just because they are anticipating that their results will be widely used and cited, then that they may consequently be highly rewarded for their efforts.

[10]See Debackere and Rappa (1994), and Rappa and Debackere (1993) for systematic explorations of problem

is *pioneering* which is obtained by computing the number of problems already solved so far in the research areas considered. The second variable is the level of *specialization* of research derived by computing the research areas' distance to the most general area (the root). The third variable relates to the *audience* of research areas. It is to be understood as the expected volume of further research that may refer to a scientific contribution appearing in a given area. Since scientific papers tend to cite papers belonging to the same or to connected areas, we will assume that the audience of a research area is linearly increasing with the size of the associated sub-field. Our tree-like representation allows us to simply define the size of an area sub-field as the size of its associated *sub-tree*[11].

Lastly, what we needed is an *instantaneous aggregation procedure*, so as to infer collective outcomes from individual choices (Kirman, 1992): we introduce a simple probabilistic rule which gives the instantaneous allocation of efforts over the research areas given the relative incentives, which in turn generate the effective 'arrival' of knowledge at each period of time. This rule is also designed to capture the idea that the concentration of competitive pressure on the most rewarding areas may also influence the dynamics[12]. Altogether, one obtains a stochastic process that inter-temporally aggregates scientists' efforts and leads to the growth of the discipline (the tree). This original process is denoted by $\left\{T^t \left| t = 1, ..., \tau \right.\right\}$. Since it is based on dynamic growing trees, this process has some common features with the *Bienaymé-Galton-Watson branching process* in applied mathematics (Kulkarni, 1995), the growing literature on tree indexed processes in Probability (Lyons and Peres, 2001) and the *avalanches* literature in Physics (for an application to economics, see Plouraboue et al., 1998)[13]. Since this is a process leading to complex dynamics, its properties are analyzed using standard *Monte Carlo experiments*. Our main results are that the process exhibits path dependency (David, 1985)

choice by recording, over a long period, scientists' entry choices in an emerging field and by surveying their motivations.

[11] Here lies one important justifications for the tree structure: focusing on incentives (problem choice) the tree-like structure captures most of the dynamics principles governing research effort allocations and thus the growth of knowledge. We are indebted to a referee for having strongly encouraged us to clarify this point.

[12] This idea has been introduced in Merton (1957) and Merton and Lewis (1971), surveyed by Zuckerman (1978) and some way discussed in the model of Dasgupta and Maskin (1987).

[13] The process of research agenda determination within science, since it is marked by sequential interdependent decisions as today's decisions to investigate a given research area will obviously be influenced by former decisions, has clearly something in common with the *information cascades* litterature (Bikhandani et al, 1991; Barnejee, 1990).

especially with regard to the level of specialization of knowledge: disciplines that are more specialized have a higher chance to become even more specialized. We also find that there is a decline in the generation of new research areas over time which, as we will show, can be compensated by increasing the rewards for performing pioneering research. We also study how the outcoming disciplines are shaped through tuning the various typical incentives of the Open Science rewarding process. Finally, an extension of the model enables us to show that a high *intertemporal correlation of choices* leads the creation behavior to collapse through a *phase transition-like phenomenon.*

The paper is organized as follows. The model is developed in the next section. The third section is dedicated to the study of the generic properties of the process, while the fourth section studies parameters effects on the dynamics and discuss the characteristics of the outcoming disciplines. The fifth section introduces intertemporal correlation and analyzes its consequences. The last section concludes.

2 The model

We first define the disciplines, seen as more or less improved research areas organized as nodes in a tree-like web. Next we discuss and show how these disciplines generate a set of problems that can be handled given the present state of knowledge. Thirdly we introduce an expected reward function which provides the incentives associated with solving each of the available problems. Finally we present the probabilistic function implementing scientists' choices at each period and thus the advances of scientific knowledge.

2.1 Scientific disciplines

At each period t of the discrete time, let a scientific discipline be described as an undirected graph G^t. A graph is formally defined as a double set: $G^t \equiv \left\{V^t, E^t\right\}$, where V^t is the set of the nodes (i) (vertices) of the graph and E^t is the set of the edges (ij) of the graph. Any edge $(ij) \in E^t$ is a link between the two nodes (i) and (j) belonging to the set of nodes: $(i), (j) \in V^t$.

We represent scientific disciplines as 'knowledge trees' where nodes designate different but interdependent research areas. Substantially, the root corresponds to the most general research area, while 'lower' nodes correspond to more specialized areas of knowledge. Since we consider trees, the extra properties follow:

- the graph is connected, i.e. there is a path relating each pair of nodes of the graph : $\forall\ (i)$ and $(j) \in V^t$, $\exists$ a path $\{(i,l),(l,u)....,(k,j)\} \subset E^t$;

- the graph is minimally connected: there is only one path between any two nodes of the graph (there is no cycle in the graph). As a consequence, we have a direct correspondence between the cardinal of V^t and the cardinal of E^t : $\text{card}\{E^t\} = \text{card}\{V^t\} - 1$.

Moreover, a tree is also a planar graph: it can always be represented on a plan without any crossing between edges. Since there is one and only one path between any two nodes, a tree has also an unambiguous geodesic distance which is a counting function of the number of edges of the path connecting two nodes. Thus $d(i,j)$, the distance between the nodes (i) and (j), is given by: $d(i,j) \equiv \text{card}\{(i,l),(l,y)....,(k,j)\}$. The root of the graph is a specific node denoted (1). The distance to the root $d(i,1)$ can simply be denoted d_i. Thus $d_1 = 0$. The root is assumed to represent the research area with the highest possible level of generality. The distance to the root then expresses the level of specialization of each research area. Let $G_j^t = \left\{V_j^t, E_j^t\right\}$ denote the sub-tree of G^t ($G_j^t \subset G^t$) the root of which is node (j). G_j^t is said to be a sub-field of the discipline G^t associated to area (j). Thus, we have by definition $G_1^t = G^t$. The size of the sub-discipline G_j^t is given by s_j^t, the number of research areas of G_j^t: $s_j^t \equiv card\left\{V_j^t\right\}$.

The tree structure (having also specified that (1) is the root of the tree) allows us to clearly define the following 'father' operator, $f_{G^t}(\cdot) : V^{t*} \to V^t$, where V^{t*} is the set of nodes of the graph but the root: $V^{t*} = V^t \ominus (1)$ [14], which gives the 'father' of any of the nodes of V^{t*}. Formally we have $(i) = f_{G^t}(j)$ if and only if both $(ij) \in E^t$ and $d_i = d_j + 1$. The inverse function: $f_{G^t}^{-1}(\cdot) : V^t \to V^{t*}$, tells us that (j) is in the childhood of (i) if $(j) \in C_i^t = f_{V^t}^{-1}(i)$.

Finally, let define the vector $\Phi^t = \left(\phi_i^t\right)_{i \in V^t}$ which assigns to each node of the graph $(i) \in V^t$,

[14] The symbol $\ominus$ simply indicates that the node considered is removed from the set. By $\oplus$, it is meant that the element is added to the set.

a non null integer $\left(\phi_i^t\right) \in \mathbb{N}^*, \forall i \in V^t$ denoting the robustness of knowledge (or its improvement level) attained within the research area (i) (thus $\Phi^t \in (\mathbb{N}^*)^{\text{card}\left\{V^t\right\}}$). It is defined as the number of problems solved in the research area considered.

Together, the set of nodes, the set of edges, and the improvement levels define the state of the discipline (the system) at each period, i.e. the knowledge accumulated so far and its structure. It is denoted by $T^t \equiv \left(V^t, E^t, \Phi^t\right)$

2.2 Available research agendas and problems generation

Let us now define the set of attainable problems given the present development of the discipline, i.e. the set of potential research agendas that the present state of the system allows to tackle. As it has been said before, one cannot reasonably assume that the set of problems is fixed and *ex ante* specified, but rather generated and even shaped by the advance of knowledge. In order to capture that feature, we consider that scientists can either i) improve knowledge in some already existing research area, or ii) create a new research area from an already existing one. In other words, the problems that may be solved are either the next problem to hand in already improved research areas or the first problem to be solved of a research area connected to any already improved one.

In this last respect, we further consider that a new 'virtual' node is associated with each existing one: that is, we consider that scientists are able to create a new and more specialized research area 'from' any existing one. Let W^t be this set of nodes that can be created at period t: each (j) in W^t is simply a *leaf* from each node $(i) \in V^t$, the improvement level of which is still null: $\forall (j) \in W^t : \phi_j^t = 0$. Obviously, we have: card$\left\{V^t\right\}$ =card$\left\{W^t\right\}$ and $V^t \cap W^t = \emptyset$.

Let us now define the set O^t such as: $O^t = V^t \cup W^t$. It is the set of opportunities, that is the set of research areas that contain a handleable problem at date t (whether the research areas have already been improved or not). We also need to define the graph that covers the opportunities, that is $\Lambda^t \equiv \left\{O^t, F^t\right\}$ such that its set of edges F^t is the union of the set edges E^t (of graph G^t) and of the set of edges say I^t that is binding each node of V^t to one distinct node of $W^t : F^t \equiv E^t \cup I^t$. The definition of the tree-like graph Λ^t, having (1) as its root, allows us to rigorously define the set W^t, which is formally such that both $i)$ $\forall (j) \in W^t, \exists$ one and only one $(i) \in V^t$, such that $(i) = f_{\Lambda^t}(j)$; and $ii)$ $\forall (i) \in V^t$, $\exists$ one and only one $(j) \in W^t$ such that $(j) = f_{\Lambda^t}^{-1}(i)$.

We assume that the state of the system simply evolves by selecting one node among the set of opportunities at each period. The selected node at period t is denoted by $\left(o^t\right) \in O^t$. Formally, when a node is selected, the two following events transform the state of the system:

$i)$

$$\begin{cases} \text{if} \quad \left(o^t\right) = (i) \in V^t \quad \text{then} \quad V^{t+1} = V^t; W^{t+1} = W^t \\ \text{if} \quad \left(o^t\right) = (l) \in W^t \quad \text{then} \\ \qquad V^{t+1} = V^t \oplus (l)\, ; W^{t+1} = W^t \ominus (l) \oplus (k) \oplus (h) \\ \qquad E^{t+1} = E^{t+1} \oplus (ul)\, ; I^{t+1} = I^t \ominus (ul) \oplus (lk) \oplus (uh) \end{cases} \tag{1a}$$

$$ii) \quad \phi_{o^t}^{t+1} = \phi_{o^t}^{t} + 1 \tag{1b}$$

$(1a)$ says that if the chosen research area is an already improved one ($(i) \in V^t$), the set of nodes and potential ones are left unchanged. If the chosen node is a new one ($(l) \in W^t$) it is then added to the set of already improved nodes, while it is withdrawn from the set of potential ones. In the meantime, two new elements are added to that set. The first one (k) is the new research area that may now be attained because of the initial improvement of (l): $(k) = f_{O^t}^{-1}(l)$. The second one, (h) , replaces (l) within the set of potential areas. Formally, if node (u) was such that $(u) = f_{O^t}(l)$, then (h) is such that: $(h) \in C_u^{t+1} = f_{O^t}^{-1}(u)$ with $(h) \neq (l)$. This feature of the model, traduces an implicit assumption of the model, namely that there is no scarcity in potential research areas. Each new investigation solving a problem generates new questions both inside its research area and in new areas.

$(1b)$ says that the level of improvement of the chosen research area is improved by one unitary increment. One can thus verify that the level of improvement of each node is equal to the sum of all the problems solved in the past within that research area: $\phi_i^t = \sum_{\tau=1,\dots,t} \Gamma(i, o^\tau)$ with $\Gamma(\cdot,\cdot)$ defined such that: $\Gamma(x, y) = 1$ if $x = y$ and 0 otherwise.

An example of such a process is represented in Figure 1 when starting with only one node at period 1 (the root). The numbers on each node are the levels of improvement ϕ_i^t (number of problems solved there), and the dotted lines represent possible areas that may be created i.e. virtual nodes. When one denotes each node by an integer which is increased by a unitary increment at each arrival, then the evolution of the tree described in Figure 1 can be formally written as in Table 1.

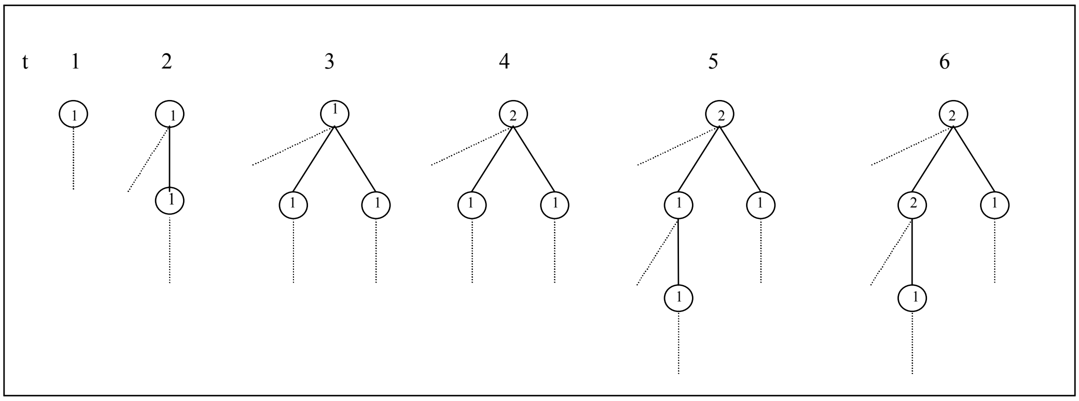


Figure 1. *An example for the evolution of a knowledge tree over the first 6 periods of time.*

t	V_t	E_t	Φ_t	W^t	O_t	o_t
1	$\{1\}$	$\emptyset$	(1)	$\{2\}$	$\{1,2\}$	$\{2\}$
2	$\{1,2\}$	$\{(1,2)\}$	$(1,1)$	$\{3,4\}$	$\{1,2,3,4\}$	$\{3\}$
3	$\{1,2,3\}$	$\{(1,2),(1,3)\}$	$(1,1,1)$	$\{4,5,6\}$	$\{1,2,3,4,5,6\}$	$\{1\}$
4	$\{1,2,3\}$	$\{(1,2),(1,3)\}$	$(2,1,1)$	$\{4,5,6\}$	$\{1,2,3,4,5,6\}$	$\{4\}$
5	$\{1,2,3,4\}$	$\{(1,2),(1,3)\,(3,4)\}$	$(2,1,1,1)$	$\{5,6,7,8\}$	$\{1,2,3,4,5,6,7,8\}$	$\{2\}$
6	$\{1,2,3,4\}$	$\{(1,2),(1,3)\,(3,4)\}$	$(2,2,1,1)$	$\{5,6,7,8\}$	$\{1,2,3,4,5,6,7,8\}$	$-$

Table 1. *The evolution of the tree like graph described in Figure 1.*

2.3 Incentives and motivations for problem choice

We now tackle the question of the incentives that may lead researchers to select projects among the available ones, that is the way the reward system in science defines an implicit payoff function for problem choice at each period. For that purpose we introduce a reward function $\omega(\cdot,\cdot,\cdot)$ that associates a given expected reward ω_o^t to each possible problem $(o) \in O^t$ at any given period of time t. This expected reward is assumed to be given by:

$$\omega_o^t = \omega\left(\phi_o^t, d_o, s_o^t\right) \tag{2}$$

ϕ_o^t has been defined as the level of robustness of the research area, *i.e.* the sum of past accumulation of knowledge on (o). When (o) corresponds to a new research area, then still no knowledge has been accumulated and thus $\phi_o^t = 0$. d_o which is the distance to the root (let us recall that $d_o \equiv d(o, 1)$) indicates the level of specialization of the research area considered (o). When d_o is low, the research area is general while if d_o is high, (o) tends to be specialized[15]. s_o^t is the size of the sub-discipline G_o^t associated with the research area (o). If (o) is a new research area then obviously, $s_o^t = 0$. s_o^t is a proxy for the "audience" associated with research areas (o).

We propose a simple Cobb Douglas specification for (2) as follows:

$$\omega_o^t = \left(1 + \phi_o^t\right)^\gamma (1 + d_o)^\lambda \left(1 + s_o^t\right)^\delta \tag{3}$$

where $\left(1 + \phi_o^t\right)^\gamma$ (with γ such that $\gamma < 0$) stands for a measure of the relative rewarding of problems in academic publications depending on their novelty. The higher ϕ_o^t the less original the next problem solved there and therefore the lower the associated reputation gain (because $\gamma < 0$). The reward of the nth contribution to a given research area is only a fraction of the first tuned by γ: when γ is small, $\left(1 + \phi_o^t\right)^\gamma$ tends to highly decrease when ϕ_o^t increases (especially when ϕ_o^t is still small). On the contrary, when γ is close to 0, the decreasing slope is lower. Thus when γ is close to 0, the first problems solved within a given research area tend to be relatively as rewarded as the later ones, while when γ is much lower than 0, pioneers will be much more rewarded than later contributors.

The second component of expression (3), namely $(1 + d_o)^\lambda$, stands here for a measure of the relative rewarding for producing knowledge in a given area depending on its generality/specialization. This parameter is typically influenced by the role of experiments, but also by the difficulty to perform valid research either because of its inherent complexity or because access to high-level equipments is needed. Parameter λ is assumed to be positive ($\lambda > 0$), which implies that the reward is always higher for specialized areas. This feature may appear counter-intuitive at first glance. Nevertheless, the more general a research area, the more difficult it is

[15] It is a specific and constant attribute of the research area considered, independent of the time period considered. Thus the time period superscripts have been removed.

to perform research there and the lower the expected reward of a standardized unit of effort. λ controls the relative preference/rewarding for specialized research areas.

The last component of the expected reward function $\left(1+s_o^t\right)^\delta$ measures the impact of the audience of research areas on the expected rewards. When the associated sub-discipline is big, i.e. is composed of a large number of research areas, then the research results that may appear in such a research area are more likely to be frequently cited. The parameter δ (such as $\delta > 0$) expresses the relative importance of that factor in the reward system, that is the strength of the citation system (or even the 'vertical' integration of the discipline). When $\delta \to 0$, the size of the sub-tree of area (o) tends to have no more any effect on the expected rewards ω_o^t. The more δ increases, the more the audience $\left(1+s_o^t\right)^\delta$ becomes an important element of the rewards (s_o^t becomes critical).

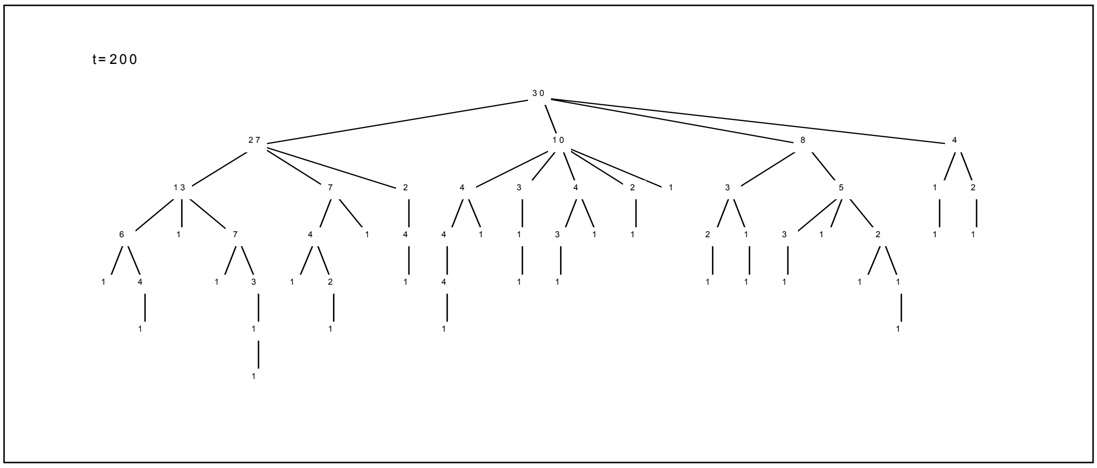
2.4 Sequential discrete choices and the stochastic process

As we emphasized before, science has a specific incentive structure which influences the determination process of research agenda by rational scientists and therefore collective outcomes. Let us assume that the problem associated with each opportunity is selected and solved with the following probability:

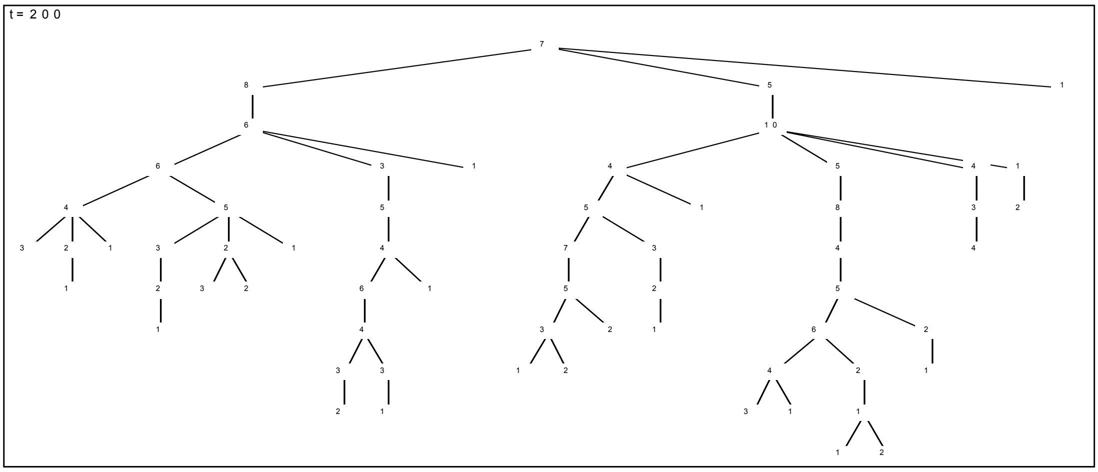
$$p_o^t \equiv \frac{\left(\omega_o^t\right)^\alpha}{\sum_{j \in O^t}\left(\omega_j^t\right)^\alpha} \qquad (4)$$

with $\sum_{o \in O^t} p_o^t = 1$, $\forall t$ and α ($\alpha > 0$). α stands for the concentration of scientists' attention on the most rewarding areas for a given allocation of incentives over the handleable problems. For instance, when $\alpha < 1$, the less rewarding research areas tend to be selected 'more frequently'. On the contrary, when $\alpha > 1$, the more rewarding areas tend to be selected more than proportionally to their expected rewards, while less rewarding areas tend to be chosen less than proportionally. As $\alpha \to \infty$ the node associated with the highest reward is almost surely chosen, and the system becomes 'quasi-deterministic'. On the contrary, when $\alpha \to 0$, each possible opportunity is chosen with the same unique probability: $p_o^t \to p^t, \forall t$ and $(o) \in O^t$.

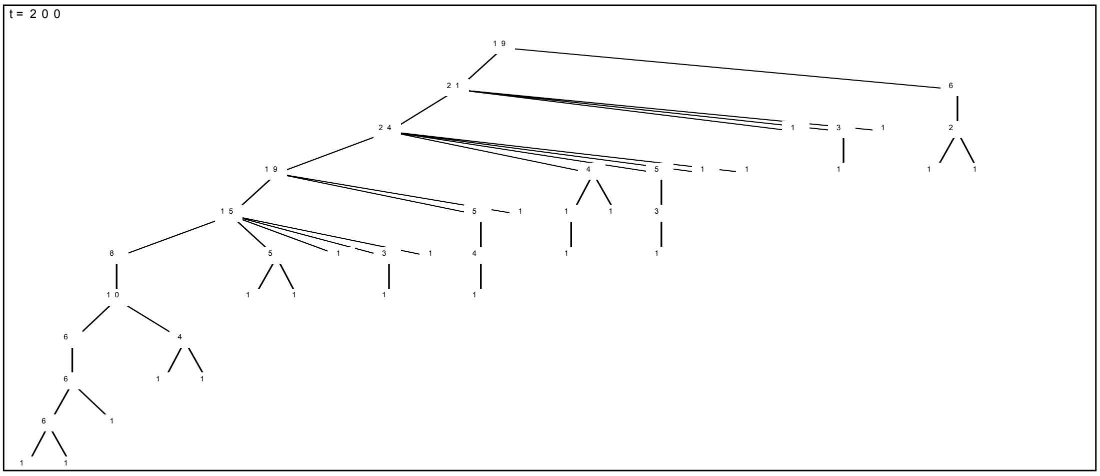
Introducing equation (3) in equation (4) gives us the probability that any research area (whether it is an existing one or a potentially created one) is improved at period t:



a



b



c

Figure 2. *Typical grown trees (200 periods).*

$$p_o^t = \frac{\left(\left(1+\phi_o^t\right)^{\alpha}(1+d_o)^{\alpha}\left(1+s_o^t\right)^{\delta}\right)^{\alpha}}{\sum_{j\in O^t}\left(\left(1+\phi_j^t\right)^{\gamma}(1+d_j)^{\lambda}\left(1+s_j^t\right)^{\delta}\right)^{\alpha}} \qquad (5)$$

This closes the description of the stochastic process of knowledge generation within disciplines. Formally, we obtain a stochastic discrete time infinite space state process that we denote by $\left\{T^t\,|t=1,...,\tau\right\}$ or conversely $\left\{\left(V^t,E^t,\Phi^t\right)|t=1,...,\tau\right\}$ since it describes the evolution through time of the two sets V^t, and E^t, and the vector Φ^t. Figure 2 shows three dynamically grown trees that have resulted from different numerical experiments realized with such a stochastic process.

Together, the parameters $(\lambda,\delta,\gamma,\alpha)$ which appear in equation (5) are said to characterize an academic community and its associated reward system. α stands for the problem choice practice given the incentives. λ, δ, and γ give the relative weighting of the various types of incentives whether this weighting comes from the motivations of the agents or from the effective rewarding. In practice both may be very close since academic communities are, to a large extent, self-organized in the sense that agents are both the rewarded and the rewarding agents.

3 Generic properties of the process

We now turn to an exploration of the generic properties of the system, that is the behavior of the dynamic process through time and its limit behavior, while the characteristics of the drawn trees depending on parameters values are described in the next section. As it has been said above, the system $\left\{T^t\,|t=1,...,\tau\right\}$ is a quite complex one which naturally leads to complex dynamics. To make that point clear, let us consider the following. From equation (5), it can easily be demonstrated that: $\forall t,\forall i\in O^t:\frac{\partial p_i}{\partial d_i}>0;\ \frac{\partial p_i}{\partial s_i}>0;\ \frac{\partial p_i}{\partial \phi_i}<0$; under the assumptions made previously about the parameters value spaces ($\gamma<0,\delta>0;\lambda>0$ and with $\alpha>0$). This implies that, all things equal, more specialized research areas, new research areas and research areas with larger audience are more attractive to scientists' choices. The problem is that these variables are dynamically correlated because the specialization and the audience variables are often opposite and because the robustness levels (ϕ_i) act as a 'crowding out' variable which tends to re-allocate incentives through time over the population of research areas (while it also regulates partly the creation of new ones). In order to analyze the complex behavior of the system, we mostly rely on Monte-Carlo simulation experiments. All experiments that will be presented in the remaining of the paper start with a tree reduced to its root at period one ($V^1 = \{(1)\}; E^1 = \emptyset; \Phi^1 = (1)$).

Among the various features of the system's behavior that one may wish to consider, we are particularly interested in the evolution of the two following ones: the generation of research areas (vs. the improvement of existing ones), and the specialization of scientific knowledge (that is how far from the root the problems solved are located). The two sub-sections below tackle these issues successively.

3.1 The inexorable decline of research areas generation

To analyze collective outcomes, it is useful to define some aggregate measures. Let us define the *creation index* as the number of nodes created over the number of possible node creations or else the number of periods, i.e.:

$$\varsigma^t \equiv \frac{1}{t}\ \mathrm{card}\{V^t\} \qquad (6)$$

Obviously, this index is such that $0 \leq \varsigma^t \leq 1$, because there is at the most one creation per period. ς^t may also be called the *exploration index* because it capture the past ability of the community to explore new scientific paths.

One has to see that there is a real trade-off between creation/exploration and robustness, because any effort which is not dedicated to the creation of a research area is dedicated to the improvement of an existing one. Thus the average improvement of research areas given by: $\bar{\phi}^t \equiv \frac{1}{\mathrm{card}\{V^t\}} \sum_{(i)\in V^t} \phi_i^t$; is the inverse of the creation index: $\bar{\phi}^t = 1/\varsigma^t$; just because we have: $\forall t, \sum_{(i)\in V^t} \phi_i^t = t$.

Figure 3 presents ten identical runs of the dynamic process, recording at each period the creation index ς^t, the parameters being arbitrarily fixed. Many other numerical experiments with different values of the parameters have been run, all showing that *creation decreases with time.* We have also run experiments over very long periods showing that the creation index was still decreasing after 10,000 periods. A unique but important exception arises when $\alpha = 0$ that is when the evolution probabilities are equal at each period. In such a situation, it can easily be demonstrated that the creation probability is equal to 1/2 at any period.

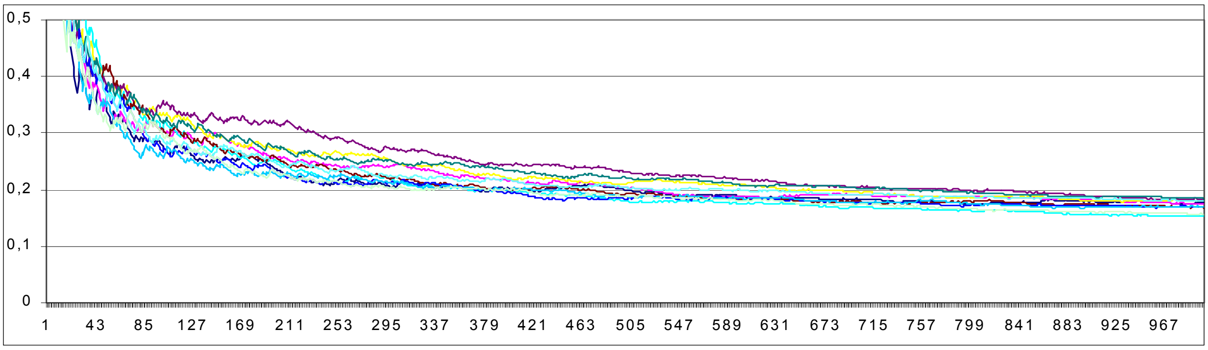


Figure 3. *Time series of the creation index ς^t over 1000 periods. 10 identical runs with:* $\delta = \lambda = -\gamma = 1$, $\alpha = 4$

This general result, is a simple but not so obvious consequence of the fact that knowledge improvements are rewarded by the attention of other scientists, often measured (even if very imperfectly) by counting citations received: the more a scientific discipline grows, the higher the incentives for improving knowledge in existing research areas just because their audience (sub-tree size) is relatively increasing as compared to incentives for exploring new paths.

3.2 General vs. specialized disciplines: a path-dependent outcome

Let us define the average generality of knowledge index, as the sum of the robustness levels of all research areas, weighted by their level of generality (the inverse of the specialization level: $1/d_i$), over the sum of all the robustness levels, as follows:

$$\tilde{\phi}^t \equiv \frac{\sum_{(i)\in V^t}\left(\phi_i^t/d_i\right)}{\sum_{(i)\in V^t}\phi_i^t} \qquad (7)$$

This index gives the average generality of each problem solved so far, which we can say is the average level of generality of knowledge within the discipline considered. Like the creation index, this index is such that: $0 \leq \tilde{\phi}^t \leq 1$, with $\tilde{\phi}^t$ close to 1 (to 0) when problems solved have been highly general (specialized).

Figure 4 presents 10 identical numerical runs of the process recording at each period the evolution of the knowledge specialization through the index $\tilde{\phi}^t$, all parameters being fixed. It startingly shows that this is decreasing through time. This feature is robust to modifications of the parameters values. The explanation for such a result is quite obvious because, in the model, it is precisely the past exploration behaviors at specialized levels that allow agents to solve even more specialized problems.

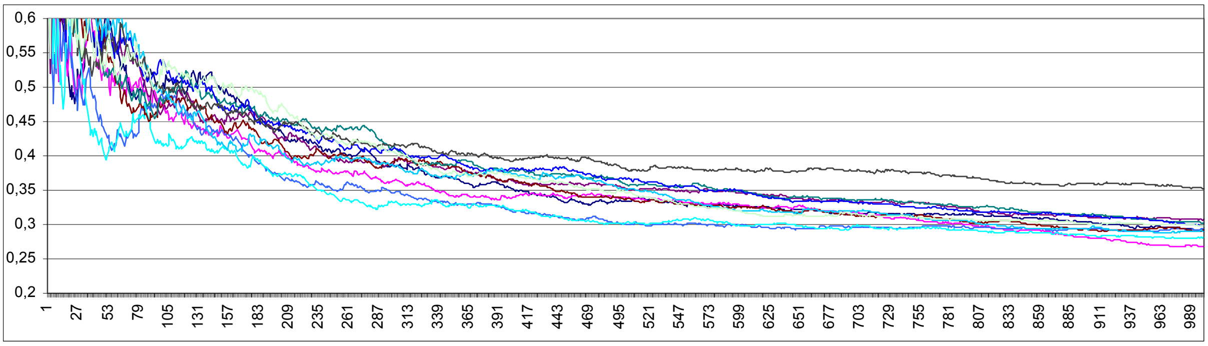


Figure 4. *Time series for the specialization of knowledge* $\tilde{\phi}^t$. *Ten numerical experiments realized over 1000 periods, with parameters fixed as follows:* $\delta = \lambda = \alpha = 1, \gamma = -1$.

The other insight that time series may provide has to do with the path dependency property, which in dynamic stochastic systems theory simply means that the transitory states of the system (state variables) determine its limit behavior. Practically, if we find that there is increasing differences between identical runs of the system (same initial states and same parameters values), thus path dependency is said to occur. If differences tend to diminish, the time series tend towards a common and unique limit, thus the system is simply auto-regressive and does not exhibit the path dependency property.

To know which phenomenon occurs, we computed and recorded the variation coefficient (variance/mean) of the generality index $\tilde{\phi}^t$ of ten identical runs of the process. As the parameter

α is likely to be critical in such respect we have done the experiment for different values of α. In order to see potential qualitative differences with the creation index, we computed the variation coefficient for creation index ς^t. The results obtained for both ς^t and $\tilde{\phi}^t$ are presented in the graphs *a* and *b* of Figure 5.

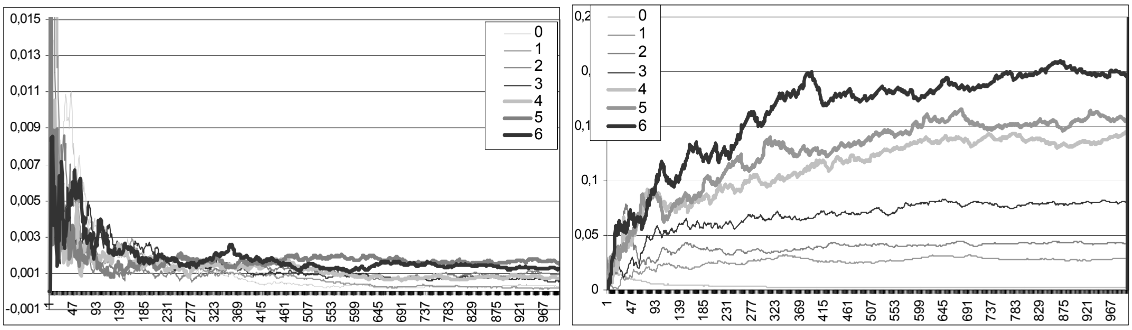
a  b

Figure 5. *Evolution, over 1000 periods, of the variation coefficient of both the creation index (graph a) and the generality index (graph b) from ten identical runs of the process, computed for different values of* α *(*$\alpha = 0, 1, ..., 6$*). Numerical experiments realized with fixed values of the other parameters:* $\delta = \lambda = -\gamma = 1$*.*

As one can observe in comparing graphs *a* and *b* of Figure 5, one gets very contrasted results between the path dependency of the system while focusing either on its creation behavior or on its generality[16]. The variation coefficient of the creation index is clearly very small and decreasing toward zero as time goes (whatever the value of its parameter α). Thus clearly, the system does not exhibit a path dependency property when looking at the creation of research areas. The opposite occurs when one looks at the generality index. Indeed, one can observe in graph *b* that the variation coefficient of $\tilde{\phi}^t$ increases with time, the slope being positively influenced by parameter α. Thus the initial events - the first choices of where to solve problems - can durably influence knowledge production. As a consequence, the dedication of scientific disciplines toward applied or more general knowledge can be influenced not only by their endogenous characteristics but also by the very history of disciplines and the path they toke at

[16]Recalling that both indexes are comprised between zero and one.

their beginning. Early pioneers of new scientific disciplines are therefore responsible for shaping them and giving them a more or less 'applied' turn.

4 Incentives, motivations and the outcoming disciplines: parameters effects study

A given combination of parameters values provides the exogenously defined relative weighting of the various typical features of Open Science rewarding. As indicated above it is giving both the specific motivations of the scientific community and the effective rewarding practices. In the following we first conduct a dedicated analysis of parameters effects on the outcoming disciplines through computing their two indexes defined in equations (6) and (7) after a given long period of time. Next we discuss the generic features of the typical disciplines generated by opposed values of parameters.

4.1 Parameters study

To get clearer results, we present several simulations according to the following protocol. First, set $\alpha = 1$. Then each time one of the three parameters (δ, λ, γ) is fixed while the two others vary. For each couple of values of the two tuned parameters, we compute the values of the indices obtained after 1000 periods. From now on, we write ς° for ς^{1000} and $\tilde{\phi}^{\circ}$ for $\tilde{\phi}^{1000}$. Every single point in the graphs presented here (Figure 5) then corresponds to the values of one index computed for a singular grown tree (thus 400 trees were generated for each simulation experiment). These 2×2 simulation process is used because it allows us to assess the robustness of the parameter modification effects for various values of the other parameters. For controling purposes, the same experiments have been reproduced for different values of the fixed parameters. Moreover, the whole experiment has been reproduced with different values of α. As a first global result, we found that when α increases the slopes in the indexes due to modifications of the other parameters become steeper. Thus the effects discussed below are valid under different values of the community's sensitivity to the academic incentives tuned through parameter α.

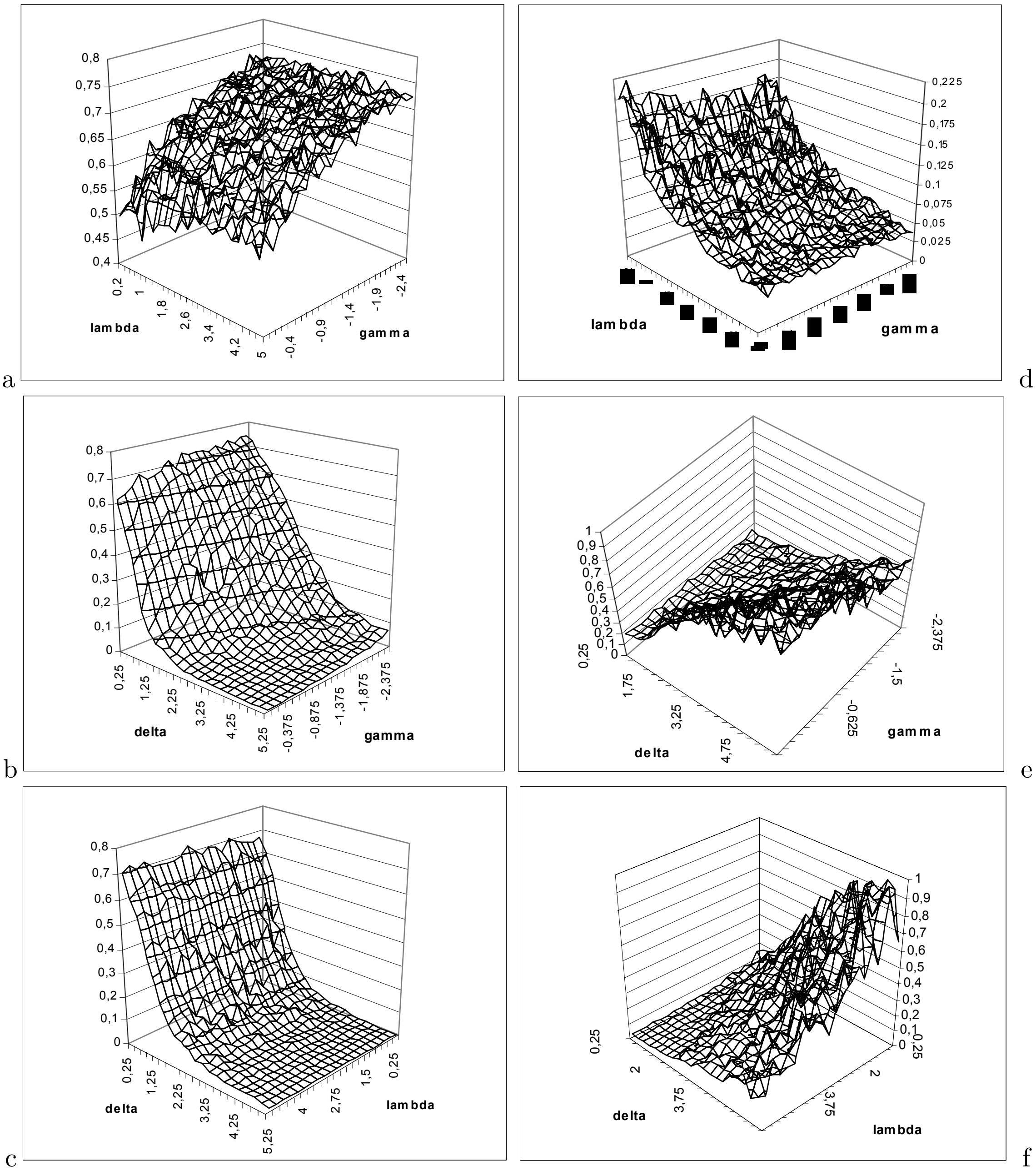


Figure 5. *Creation index ς° (left: graphs a, b, and c) and generality index $\tilde{\phi}^{\circ}$ (right: graphs d, e and f). Graphs a and d: λ and γ vary, $\delta = 1$. Graphs b and e: δ and γ vary, $\lambda = 1$. Graphs c and f: δ and λ vary, $\gamma = -1$. In all experiments, results are observed after 1000 periods and with $\alpha = 1$.*

Table 2 summarizes the main results obtained.

	δ	γ	λ
ς°	$\searrow$	$\searrow$	$\longrightarrow$
$\tilde{\phi}^\circ$	$\nearrow$	$\nearrow$	$\searrow$

Table 2. *Effects of the parameters (δ, γ, λ) on the indexes (ς°, $\tilde{\phi}^\circ$). A positive effect is denoted by $\nearrow$, a negative by $\searrow$, and no effects by $\longrightarrow$.*

Both a higher δ and a γ closer to 0[17] decrease creation: looking for citation rents, academic scientists tend to prefer improving knowledge within existing areas rather than exploring new ones, both when the size of the audience (an important sub-tree) counts much and when rewards for performing pioneering research are lower. Higher δ tends to favor research performed at a high level of generality because of the following: research areas which are more general are often (but not systematically) the ones that have the strongest audiences, i.e. research performed there is likely to be cited by research performed in more specialized but connected areas. Thus, general areas become naturally more attractive on average. In a similar fashion, since γ regulates the incentives to dedicate research efforts toward more or less improved research areas, when it increases it diminishes the creation of research areas and prevents the discipline from achieving an important level of specialization. That naturally comes from the fact that the model assumes specialization to result from successive downward exploration. Thus some pretty general disciplines may in fact be 'under-developed' trees, therefrom lower specialization may be, in certain circumstances, just a consequence of a lack of exploration behaviors. Finally, higher λ increases the attractivity of more specialized research areas, and therefore the average distance to the root.

4.2 How motivations and incentives lead to various outcoming disciplines

We have just seen how the features of the outcoming disciplines are influenced by each parameter. We now take the opposite point of view, by taking polar forms of the discipline-trees and wonder how they may be generated? Furthermore: What is their probability of

[17] Recalling that γ is negative, thus "γ increases" is similar to "γ becomes closer to zero".

occurrence because of reinforcing (or opposed) effects of the parameters? And finally, are these forms likely to be stable over time considering the generic properties of the process highlighted in the previous section?

Let us first define the four *polar forms* of disciplines coming from opposite values of the two indexes: the 'star' form is obtained when the discipline is both highly general and research areas are highly improved (high $\tilde{\phi}^\circ$, low ς°), a 'well' form is said to arise when the discipline is highly specialized and counts only few research areas (low $\tilde{\phi}^\circ$, low ς°), the 'flake' form comes from many creations at a quite general level (high $\tilde{\phi}^\circ$, high ς°), and the 'rake' form stands for both high specialization and many creations (low $\tilde{\phi}^\circ$, high ς°). These configurations are exposed in Table 3 below.

	high $\tilde{\phi}^\circ$	low $\tilde{\phi}^\circ$
low ς°	**'Star'**	**'Well'**
high ς°	**'Flake'**	**'Rake'**

Table 3. *The polar forms of the scientific disciplines.*

We have seen that δ and γ play systematically in an opposite manner. Thus, in the one hand, they can counterbalance each other, in the sense that the parameter values which lead to specialized (general) disciplines are also the same that prevent their areas to be much improved (numerous). That makes both the 'star' and the 'rake' forms quite unlikely to emerge since the specific traits of each form are generated by opposed values of the parameters. On the other hand, they may also reinforce one another. That leads to the idea that the 'flake' and the 'well' forms are quite probable ones: the factors that orient the disciplines towards being specialized (general) tend also to orient the research efforts toward being less (more) exploratory.

Nevertheles, as we have seen before, the creation of new research areas is likely to diminish through time: thus the 'flake' forms might be seen as a transitory state of the 'star' that may emerge in the long run: once the community has generated a certain number of research areas, the relative incentives for improving research areas may overbalance the exploration ones, and thus, among the existing areas, the general ones which have a larger audience, become more attractive. As a matter of fact, the 'star' form has much in common with a disciplines such as Physics, which is an old and fully integrated discipline, which has explored many specific and applied areas and where the most theoretical and general discoveries may benefit to a large audience dispersed in many specialized research areas.

On the opposite corner, the 'well' form seems to be a quite interesting and robust typical shape of scientific disciplines since specialization is a path dependent outcome and exploration tends to decrease. Such disciplines emerge because of weak incentives for performing pioneering research and weak rewarding of research due to its relative audience. The latter may come from scientific communities that would not (or weakly) take citations into account in the evaluation process as compared to publication counts. In such circumstances, the attention of scientists may follow one single (or few) line of investigation, while excluding many research opportunities in a quite 'autistic' fashion. Thus, one obtains a major 'science well' as illustrated in Figure 6.

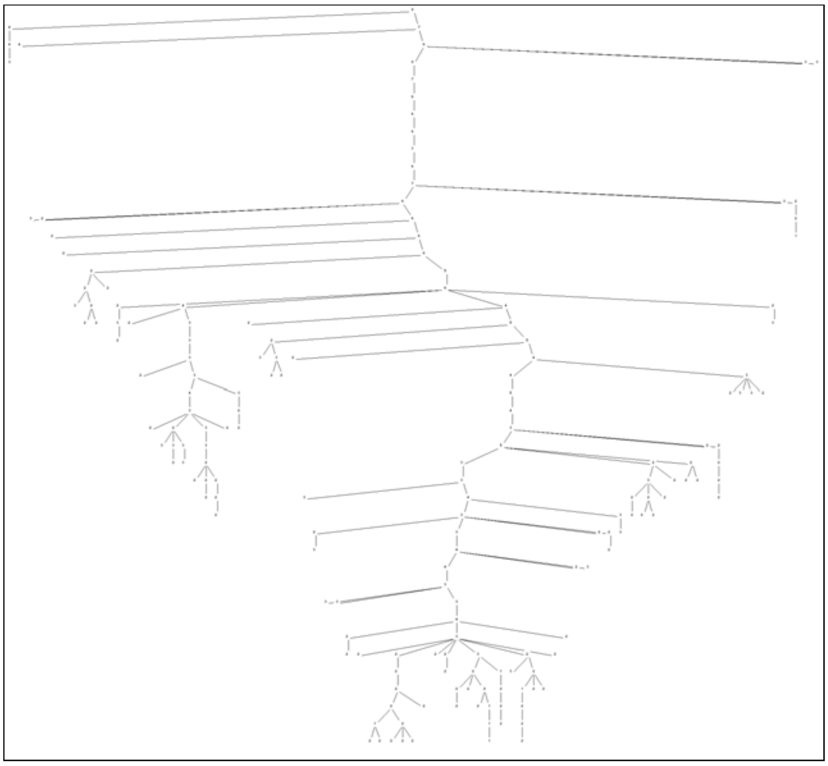


Figure 6. *A 'science well'.*

5 Extension: Intertemporal correlation of choices

In this section we introduce a new feature of the dynamic process, namely an *intertemporal correlation of choices* which stipulates that present choices are directly influenced by previous ones. Two main rationales motivate that trait. The first one has much in common with what leads scientists to say that a given research areas is 'hot'. A 'hot' research area is often one where discoveries recently occurred. It has been showed in survey-like studies of problem choice, that such factor is one of the main reasons that motivates researchers to investigate scientific domains (Rappa and Debackere, 1993). That is probably because academic agents may rationally take recent discoveries as a signal that there is still much to be done there. Moreover, there is also some anecdotal evidence that support the idea that such 'hot' areas tend to highly attract researchers' attention even if they are not precisely specialize in that area. Thus there is more reputation to be gained in such areas audience the audience of which is temporary increased going beyond the classical boundaries between sub-fields. The other justification for introducing an intertemporal correlation is due to cognitive specialization of knowledge production. Knowledge can be considered as having a local character (Stiglitz, 1987) in the sense that agents do specialize in specific areas of a discipline. Thus *cognitive mobility* is costly, and reallocation of efforts may be quite costly for agents, especially when they have invested much and when the time to go before retirement is getting shorter (Stephan, 1996). Thus the problem choices become globally sticky that is the reallocation of efforts due to recent modifications in incentives is lowered. In the following we introduce definitions and subsequent improvements of the process before presenting the results.

5.1 Definitions

Let the vector $\Psi^t \subset \mathbb{N}^m$ denote the memory of the system at period t. This vector records the research areas where the m ($\in \mathbb{N}$) last problems were solved, that is during the m last periods. Thus this parameter gives the length of the window for the systems memory. The memory vector is formally defined as: $\Psi^t = \{(o_\tau)\}_{\tau=t-m-1,\ldots,t-1}$, recalling that (o_τ) denotes the research area where a problem has been solved at period τ. This leads to a state of the system which is now completely defined by $T^t = \left(N^t, E^t, \Phi^t, \Psi^t\right)$.

Let us now denote the hotness of any research area (o) at period t by ψ_o^t, which is defined as the number of past evolutions that affected it during the last m periods. That is:

$$\psi_o^t = \frac{1}{m}\sum_{o^\tau \in \Psi^t} \Gamma(o^\tau, o) \qquad (8)$$

with $\Gamma(\cdot,\cdot)$ defined in Section 2. Thus, one may label a *cold* research area as an area which has had no problem solved during the m last periods, *i.e.* when $\psi_o^t = 0$. Similarly, a sub-discipline G_i^t is said to be cold when $\sum_{o \in V_i^t} \psi_o^t = 0$.

Now let the incentives for problem choice be modified by the new variable introduced, through the following slight modification of the expected reward function given in equation (3):

$$\omega_o^t = (1+s_o)^\delta \left(1+d_o^t\right)^\lambda \left(1+\phi_o^t\right)^\gamma \left(1+\psi_o^t\right)^\beta \qquad (9)$$

The parameter β $(\beta > 0)$ tells how much agents' choices are influenced by the hotness character of research areas

5.2 A phase transition phenomenon

In order to study the effects of areas hotness $\left(\psi_o^t\right)$ on the system's behavior, we let its associated parameter β vary. A series of numerical experiments has been realized for fixed values of the other parameters. Some results of such simulations are presented in graphs a and b of Figure 9. Therein, one may observe that the creation index decreases sharply down to zero when β increases. Moreover, the generality index appears to increase sharply for a critical value of β in a *phase transition*-like phenomena having a threshold at $\beta_c \simeq 3$.

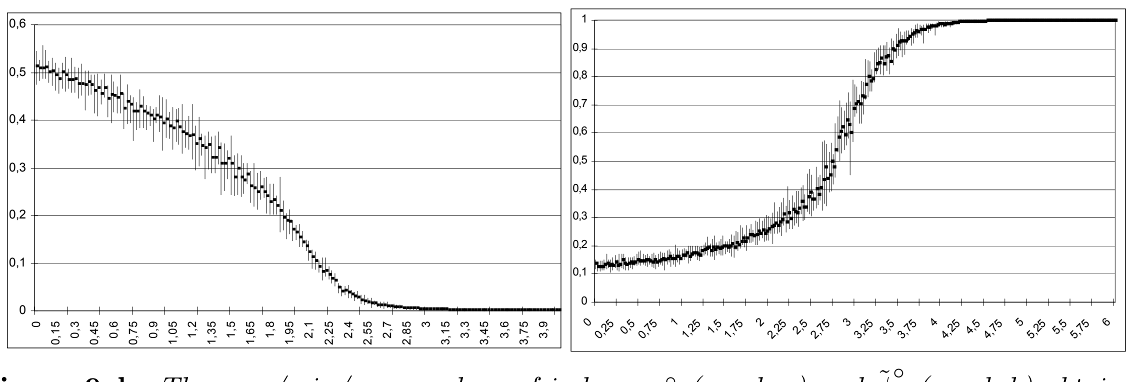
a  b

Figure 9ab. *The max/min/mean values of indexes* ς° *(graph a) and* $\tilde{\phi}^\circ$ *(graph b) obtained for different values of* β*. Other parameters:* $\alpha = \delta = \lambda = -\gamma = 1, m = 10$.

This properties were reproduced for very different values of the parameters. We also studied the possible influence of the memory size parameter (m) and found no evidence of any influence on the dynamics.

5 Conclusion

In this paper we have presented an original model of knowledge production within scientific disciplines. It is a graph theoretical model in which knowledge production is sequential. The main question tackled in the paper is how the specific incentives provided by the academic reward system influence researchers' problem choice and thus shape the stochastic process of knowledge generation within scientific disciplines. Let us sum up the main results obtained.

We first found that the process exhibits a sustained decline in the generation of new areas. This phenomenon is caused by the specific reward system in science which leads researchers to seek for others' attention. When the discipline grows, the relative rewarding of problems located in already developed fields increases: because their audience becomes larger, contributions to such domains are more likely to be cited. This is not to be seen as *a fortiori* negative: since more knowledge are likely to be produced in larger domains, present contributions there are likely to benefit to many late improvements. That is directly connected to the fact that the citation system traces and rewards knowledge spillovers. Nevertheless, this first result suggests that the rewards for performing pioneering research should be increased when the discipline grows in order to asymptotically sustain research areas generation. We next found that the stochastic process exhibits path dependency with regard to the specialization of disciplines. More specialized disciplines tend to become even more specialized through time. We found that this property is enhanced when the concentration of scientists' attention on the most rewarding areas is stronger.

In addition to these first series of results, parameter studies allowed us to highlight the possible occurrence of a quite 'autistic' dynamics leading to a 'well' form of discipline having left many research opportunities unexplored. We found that increasing the relative rewarding of pioneering research is again a key leverage parameter because, under such circumstances, it also (unexpectedly) renders general problems more attractive. We argued that, such a situation is more likely when the relative rewarding through recording citations is outweighed by publication counts. Thus, reinforcing the former mechanism may partly prevent from getting such 'science well'. Lastly, as empirical evidence suggested, we studied how an intertemporal correlation of choices may affect the dynamics. We found that the system collapses when such an effect is high. This suggests that scientific communities should be prevented from being too sticky in connection with problem choice and should preserve the collective ability to reallocate efforts over problems.

Bibliography

Arrow, K.J., 1962. Economic welfare and the allocation of resources for invention. in: Universities-National Bureau Committee for Economic Research, The Rate and Direction of Inventive Activity, 609-626. Harvard Business School Press, Cambridge.

Arrow, K.J., 1987. Reflections on the Essays. in: Feiwell, G.R. (ed.), Arrow and the Ascent of Modern Economic Theory, 685-689. Macmillan, New York.

Barnejee, A., 1990. A simple model of herd behavior. Quarterly Journal of Economics, 107, 797-818.

Bikhchandani, S., Hirshleifer, D., Welsh, I., 1991. A theory of fads, fashion, custom and cultural change as information cascades. Journal of Political Economy, 100, 992-1026.

Carayol, N., 2001. Properties and failures of open science. Essays in the economics of 'imperfect science'. Ph.D dissertation, University of Toulouse, France.

Carayol, N., 2000. Modeling creation vs. diffusion of structured knowledge, in: Ballot, G., Weisbuch G., (eds.), Applications of Simulation to Social Sciences, 353-368. Hermes Science Publications, Paris.

Cournot, A., 1861. Traité de l'enchaînement des idées fondamentales dans les sciences et dans l'histoire, Tome II. Librairie de L. Hachette, Paris (reedited in 1968, Edizioni Bizzanni, Rome).

Dalle, J.-M., 1998. Heterogeneity and rationality in stochastic interaction models. in: Cohendet, P., Stahn, H. (eds.), The Economics of Networks: Behaviors and Interactions, 123-145. Springer Verlag, Berlin.

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy, 23 (5), 487-521.

Dasgupta, P., David, P.A., 1987. Information disclosure and the economics of science and technology. in: Feiwell, G.R. (ed.), Arrow and the Ascent of Modern Economic Theory, 519-542. Macmillan, New York.

Dasgupta, P., Maskin, E., 1987. The simple economics of research portfolios. The Economic Journal, 97 (387), 581-595.

David, P.A., 1985. Clio and the economics of QWERTY. American Economic Review (Papers and Proceedings), 75, 332-337.

Debackere, K., Rappa, M. 1994. Institutional variations in problem choice and persistence among scientists in an emerging field. Research Policy, 23, 425-441.

Diestel R., 1997. Graph Theory. Springer.

Durlauf, S.N., 1997. Rational choice and the study of science. mimeo, University of Wisconsin at Madisson.

Kirman, A.P., 1992. Whom or what does the representative individual represent. Journal of Economic Perspertives, 6, 117-136.

Kulkarni, V.G., 1995. Modeling and Analysis of Stochastic Systems. Chapman & Hill, London.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago.

Lyons, R., Peres, Y., 2002. Probability on Trees and Networks. Cambridge University Press, Cambridge.

Machlup, F., 1982. Knowledge: Its Creation, Distribution and Economic Significance, Vol. II: The Branches of Learning. Princeton University Press, Princeton.

Merton, R.C., Merton, R.K., 1989. Unanticipated consequences of the reward system in science: A model of the sequencing of problem choices. Unpublished Manuscript, announced in Rationality and Society.

Merton, R.K., 1957. Priorities in scientific discovery. American Sociological Review, 22 (6), 635-659.

Merton, R.K., 1942. The normative structure of science. Journal of Legal and Political Sociology, 1, 115-126.

Merton, R.K., Lewis, R., 1971. The competitive pressure: The race for priority. Impact of Science on Society, 21 (2), 146-175.

Nelson, R.N., 1959. The simple economics of basic scientific research. Journal of Political Economy, 67, 297-306.

Peirce, C.S., 1896. The economy of research, reprinted in: C. Hartshorne and P. Weiss (eds.), 1960, Lessons from the History of Science, Collected Papers of Charles Sanders Peirce, Vol 1, 48-49. Harvard University Press, Cambridge.

Plouraboue, F., Steyer, A., Zimmermann, J-B., 1998. Learning induced criticality in consumers' adoption patterns: A neural network approach. Economics of Innovation and New Technologies, 6, 73-90.

Polanyi, M., 1962. The republic of science: Its political and economic theory. Minerva, 1 (1), 54-73.

Rappa, M.A., Debackere K., 1993. Social and cognitive influences on problem choice in R&D. American Academy of Management (Best Paper Proceedings), 347-351.

Salter, A.J., Martin, B.R., 2001. The economic benefits of publicly funded basic research: A critical review. Research Policy, 30, 509-532.

Stephan, P., 1996. The Economics of science. Journal of Economic Literature, 34, 1199-1235.

Stiglitz, J.E., 1987. Learning to learn, localized learning and technological progress. in: Dasgupta, P. and Stoneman, P. (eds.), Economic Policy and Technological Performance, 125-153. Cambridge University Press, Cambridge.

Weitzman, M.L., 1998. Recombinant growth. Quarterly Journal of Economics, 113 (2), 331-359.

Ziman, J., 1994. Knowing Everything about Nothing. Oxford University Press, Oxford.

Ziman, J., 1987. The problem of problem choice. Minerva, 25, 92-106.

Zuckerman, H.A., 1978. Theory choice and problem choice. Sociological Inquiry, 48, 65-95.