The Creative Development of Fields: Learning, Creativity, Paths, Implications

Jonathan S. Feinstein

ONLINE FIRST

Springer

Springer

CrossMark

The Creative Development of Fields: Learning, Creativity, Paths, Implications

Jonathan S. Feinstein[1]

Abstract I present a model of the creative development of a field and analysis of the model based on an extensive set of simulations. A field as defined here is a domain for human activity and engagement, including for example standard academic fields as well as practical fields like law, medicine, design, and fields of technology. In the model in this paper, the field is defined in terms of the body of knowledge and elements or products that have been created in the field up to that point. The field begins from an initial state and grows as individuals enter the field and make new contributions; its basic structure resembles a lattice. New elements are created via combining preexisting elements, based on specific rules for combinations; thus I follow much of the creativity literature in defining creativity as creating novel conceptual combinations. The heart of the model is a rational, optimizing model of individual creative development, in which individuals have as their aim maximizing the expected value of their contribution to the field. An individual selects an initial set of elements in the field to learn, then gains intuitive signals about potentially fruitful new combinations based on this learning set, selects additional elements to learn, and finally chooses a potential new element to attempt to make. If the element is viable, it is added to the field, together with any subbundle elements co-created with it. The simulation analysis reveals a rich set of empirical predictions about the development of fields through this process. A first striking find is the diversity of possible paths of development starting from a given initial state. The intuitive signals individuals receive are an important factor in generating this diversity, as signals lead individuals to attempt to make elements they might otherwise not pursue, thus shaping the development of the field in important ways. The results also reveal a high degree of

✉ Jonathan S. Feinstein
jonathan.feinstein@yale.edu

[1] Yale School of Management, P.O. Box 20-8200, New Haven, CT 06520-8200, USA

Springer

path dependence, generated as individuals build on the work of their predecessors, and interesting temporal patterns for how output in one period is linked with what occurred in the previous period.

Keywords Creativity · Innovation · Economic growth · Learning

Introduction

In this paper, I present a model of the creative development of a field. The field is defined as an explicit knowledge structure that starts from a simple initial state, then develops through the series of creative contributions made by successive individuals who enter the field. New elements are added by combining preexisting elements in new combinations. The heart of the model is a rational, optimizing model of individual creative development. Individuals have initial "seed" learning, then gain intuitive signals about potentially fruitful new combinations of elements or sub-topics in the field; their signals guide them as they choose further elements to learn and then a new element to attempt to make, basing their choices on expected value calculations. When an individual is successful in his project, the new element he creates is added to the field. The field thus grows over time.

I present an analysis of the model's implications based on extensive simulations. A striking finding is the diversity of possible paths of development starting from a given initial state, in simulations hundreds of paths even for very simple initial conditions. In addition, the results show very substantial path dependence in how a field develops, as individuals build on the work of their predecessors. The results demonstrate the importance of intuitive signals that guide individuals in generating this diversity of paths, for many paths are generated in the full model with signals that are not created in a null "clean" model in which there are no signals. The model predicts distributions for output and the number of elements created over time. I also explore the dynamics of growth, including how expected output in a period depends on the choices and outcome of the preceding period, generating implications about the time series properties of output in the development of a creative field.

The paper fits in the large literature on modeling the innovation process that drives economic growth and cultural development. The importance of innovation has been recognized from the beginnings of the modern literature on growth (Solow 1957) and in the Austrian focus on knowledge and individual initiative (Hayek 1960). Indeed, the central role of innovation in the development of industries was emphasized by Marshall in *Industry and Trade* (1919) and the importance of freedom of expression, creativity, and experimentation is the focus of John Stuart Mill's *On Liberty* (1859). In the modern theory of economic growth, the production of new ideas is central to economic development (Aghion and Howitt 1992; for an institutional, historical perspective see Mokyr 2002).

The model in this paper focuses on modeling the creative process in the context of a field of creative activity, which could be a scientific or other intellectual field, a field of technology or design, or a practice. It relates to models of searching for the best new alternative from a distribution of possibilities (Evenson and Kislev 1976; Kortum

1997; Fleming and Sorenson 2004); it also relates to Garicano's model of problem-solving (a form of creativity) in organizations (Garicano 2000). Here, creativity is modeled specifically as combining existing elements in new ways to create new elements. This approach is based on the widely accepted definition of creativity in the field of creativity studies as connecting or relating preexisting elements that have not previously been connected or related (Mednick 1962; Koestler 1964; Poincaré 1908, 1952); I discuss this definition and how it relates to other views of creativity further in "Creating New Elements." My approach is connected to the important contribution of Weitzman on recombinant growth (1998) (Feinstein 2011), although I focus more on the creative process and less on resource limits on the development of new ideas. The model formalizes and builds on the model of creative development developed by Gruber (1974), Feinstein (2006) and Cohen (2009), among others. It is a model of a field so that new ideas are generated in the context of the field (for a recent different formal model of the development of a field see Bramoullé and Saint-Paul 2010). In turn, this allows a representation of how knowledge in a field grows over time, and reveals the structure of the field in terms of how new ideas are generated based on and in relation to older ideas. Interestingly, the structure of the field resembles a lattice, thus also providing a link with the field of economic and social networks (Jackson 2008).

Explicitly modeling the creative process that drives innovation and knowledge creation is important. By formally modeling how this process works, we will be better able to understand and predict the dynamics of how economies and fields develop. This includes how human agents respond creatively to shocks, as well as how they generate new ideas endogenously within a field. The aim is not to predict the exact next idea or innovation, but rather to build models that enable us to calibrate and appreciate the range and distribution of outcomes that may arise given the current state of a field. As I show with the results of this paper, the range is in fact large, and in fact is itself highly variable for different historical paths starting from the same initial condition, with a high amount of path dependence.

A key motivation of this paper is to present a framework that links economic models of creativity and innovation with the field of knowledge representation. Knowledge representation provides a conceptual framework for describing concepts and their relationships (Sowa 1984, 2000; Wille 1992; Ganter et al. 2005). It is also useful for natural language based description (Helbig 2006); for example, Kaplan and Vakili (2013) have developed a text-based approach for linking patents. Lancaster's (1966) model of attributes is a well-known related approach in economics that also fits with the model in this paper. I develop a simple example of a knowledge representation framework of a field and use it to explore how the field develops; details are given in "Attributes and Elements" and "Creating New Elements." Specifically, I define elements in the field as strings made up out of basic "letters" or attributes. New strings are created by combining two preexisting strings according to defined rules. Success is uncertain: an element that is attempted has a probability of being viable. A viable new element has associated an output drawn from a distribution that defines its economic value; this distribution has the properties that are recognized as empirically important for the creation of innovations, specifically a long right tail so that there is a small probability of a very high value contribution.

The heart of the model is a rational, optimizing model of individual creative development. The model centers on learning, intuitive signals, and project selection choice guided by the signals. The signals pertain to subbundles of attributes that may be created when an element that embeds them is created. Thus, intuition is not necessarily about a fully defined final product, but about a "wish list" bundle of attributes, and the task is to find a way to create a new element that embeds this bundle. This captures the commonsense view of the creative process as guided by partial ("fuzzy") vision or simple insight that is then developed, filled in, and perfected. Each time period, a single individual enters the field and goes through a creative development process described in detail in "Creative Development." The model is non-stationary in that the field grows over time as new contributions are made, and the set of feasible new combinations changes and in general also grows over time. Lastly, the field has a public history that records for each individual who worked in the field in the past, his learning choices, as well as the creative project he attempted and its outcome. Probabilities about likelihoods of success in creating new elements and associated output levels are updated at the start of each period using this public history, including indirect inferences about intuitive signals individuals received based on their observed choices.

Overall, the model in this paper provides a basis for describing the dynamics of development of a field that is far more rooted in learning and rational choices than psychological models, including the Darwinian model of random variation and selection (Campbell 1960; Simonton 2003). The Darwinian model does not incorporate a rational, forward-looking learning process; it has no explicit role for either intuition or expected value calculations being made to guide choices of what to learn and attempt to combine into new elements (Gabora 2005 offers a related critique).

I analyze the model through extensive simulations. I explore several different parameter cases, for each running both the full model as well as companion runs with no signals, which I call *clean* runs. For each simulation, I generate a *masterlist* that specifies for each potential new element whether or not it is viable and its value (output) if it is viable and is produced. I run each simulation out five periods, identifying every possible path of development of the field assuming that individuals follow the optimal creative development strategy; paths differ in the set of intuitive signals individuals receive, which guides their choices about learning and which new element to attempt to create.

The results generate a set of interesting results. A striking finding is the diversity of possible paths of development starting from a given initial state. A typical simulation generates several hundred paths and a substantial number of distinct field structures (multiple paths may lead to the same structure): many simulations generate more than 100 distinct field structures through five periods. In addition, the number of distinct structures itself varies widely across simulations: Thus, there is variability in the potential variability of development of the field, depending on parameter values and outcomes early in the history of the field. Linked to the diversity of structures, there is also substantial variance in the number of new elements created and generated output. There are many fewer paths when the model is run with no intuitive signals guiding choices. Thus, an important finding is that much of the diversity of possible paths of development of the field is generated by the intuitive signals, which lead

individuals to attempt to make elements they would otherwise not attempt, in turn opening up new frontiers for future development.

A second finding is the high degree of path dependence. Comparing pairs of paths for which the element created (or attempted but not viable) in the first period is different, typically each path leads to new elements that are not created along the other path. This result fits naturally with the cumulative nature of creative development of the field, as individuals build on the work of their predecessors. A noteworthy feature, at least through five periods, is that once two paths diverge, they do not tend to reconverge, but generate substantially different sets of elements going forward.

A third set of findings concerns conditional output calculations and the dynamics of output. A basic distinction is whether a new element was successfully created or not in the preceding period—is productivity higher when a new element was successfully created? I address this question with the simulations. A second issue is whether the individual in the preceding period was guided by a signal in his choice of element to make—it stands to reason that when an individual has been guided by a signal and has been successful, this will open up new frontiers that might otherwise remain unexplored. Results in general confirm this, with conditional expected output greatest when a new element has been made in the preceding period based on a live signal. Thus, the model generates implications about the time series of output. Model results also speak to how frequently an individual attempts to build a new element using as a building block an element created in the immediately preceding period, versus an element created further back in time. I have also explored an extension in which individuals earn a royalty if an element they create is used in the future; I discuss this in the conclusion.

The remainder of the paper is organized as follows. The next section introduces basic terminology about the field, its knowledge structure, the creation of new elements, and valuation of new created elements. Section "Creative Development" describes the process of creative development of individuals working in the field, including their learning process, intuitive signals, and their optimal strategy. In "The Development of the Field" section, I describe the development of the field including its structure and how the history of the field is updated after each generation. In "Simulation Results" section, an extensive set of simulation results are presented and discussed. Section "Conclusion" concludes. An (online) Appendix (available at www.jonathanfeinstein.com) contains supplemental materials.

The Field: Elements, Creativity, Values

A field is defined to be a domain for human activity and engagement, comprising a body of knowledge and tangible elements. Knowledge may be taken broadly to include descriptive knowledge, theoretical concepts and definitions, attributions, assumptions, and theoretical statements or propositions; empirical knowledge, including facts, cases, data and examples, and heuristics; and tangible elements including materials, apparatus, and products that have been created in the field. In this paper, I use a very simplified representation of a field, in order to tie the model to a rational model of learning and creativity generation; but I believe the model can

be extended to richer contexts. At a given point in time, a field includes both widely accepted knowledge as well as opportunities for the further creation of knowledge (an important extension of the model here would allow for contested knowledge that might be challenged in new work), including new products. In reality, the boundaries of a field are typically not sharp and individuals may draw on knowledge from "outside" the field to produce new elements in a field. I do not distinguish inside and outside but rather simply assume all relevant elements for creativity generation are included in the representation. A useful extension is to allow for some parts of the representation to be inside and other parts outside the field, with implications for how individuals learn these different kinds of elements. Examples of fields include standard academic fields like the physical sciences, mathematics, the humanities and social sciences, as well as more practical fields like law, design, medicine, and fields of technology. I do not tie the model to a specific field; rather I employ a broad, conceptually abstract notation to describe the field.

There are many other elements involved in the ongoing development of field beyond the knowledge elements, including the people who are involved, institutions, resource allocation processes, and oftentimes consumers of the products created in the field. I do not focus on these additional factors in order to maintain as stream-lined a focus as possible on the main elements of the model. However, it is natural to consider extending the model, for example to consider the role of incentives on the development of the field. I have as one example explored the role of royalties on the pattern of development of the field; I discuss this extension in the conclusion. In a companion paper (Feinstein 2015), I have also explored the design of an educational curriculum for individuals who will pursue creative work in a field.

Individuals who work in a field aim to make new contributions, with the result that the field may grow and develop over time. While in some cases individuals may work in a field purely for their own enjoyment, in many cases, they do so with the aim of gaining recognition or monetary rewards. I assume that individuals gain value when they successfully produce new elements and that different elements generate different amounts of value, according to a stochastic distribution specified below. In reality of course, the overall value of an element is typically more than what accrues to the individual who produces it, but here I assume for simplicity they are the same.

Fundamentally, in terms of its knowledge structure at a point in time, the field consists of elements and relationships among elements. More complex elements are built up out of simpler elements as the field grows; I focus in particular on the creative process through which individuals working in the field create new elements. As these complex elements are created, they create links among existing elements, creating new relationships among elements. The approach follows the formal concept analysis approach (Wille 1992) and more broadly the knowledge representation framework outlined for example by Sowa (1984, 2000).

Textbook accounts of a field give a sense for the elements in a field, up to some level of complexity. An example is the well-known graduate text *Microeconomic Theory* by Mas-Colell et al. (1995). Chapter 1 is entitled "Preferences and Choice" and introduces a set of definitions about preferences and preference orderings and propositions about these elements. Chapter 2, "Consumer Choice," applies material from Chapter 1 to decisions about consumption, made subject to a budget constraint;

it develops definitions of demand functions and comparative statics and the weak axiom of revealed preference. Chapter 3, "Classical Demand Theory," includes utility maximization, expenditure minimization and duality, integrability, and the strong axiom of revealed preference. In total, the book contains 23 chapters divided into 5 parts.[1] Clearly, a knowledge structure for a well-developed field is very rich. In this paper, the knowledge structure will be quite simple relative to these applications, so as to allow focus on development of a formal model of how a field develops.

Attributes and Elements

The most primitive units in the field are *letters*. There are N letters; a_i denotes letter i. We assume letters cannot be subdivided further.[2] Letters can be thought of as primary attributes as one interpretation of the model.[3]

Elements in the field are ordered strings of letters. Order matters: thus $a_1 - a_2 - a_3$ is taken to be a distinct element from $a_2 - a_1 - a_3$. However, which end of a string is given first does not matter: $a_1 - a_2 - a_3$ is the same as $a_3 - a_2 - a_1$. New elements are created by combining two parent strings. There are rules for how strings are created and thus which strings can be created given the current state of the field. These rules are described below. Not all strings are viable; viability is discussed in "Success Rate and Value Distributions." When a string is viable, it has a value drawn from a distribution—this is also specified in "Success Rate and Value Distributions." The overall value of a string is the sum of the value of the main string and the values of any subbundles co-created with the main string, also discussed below.

Beyond the direct interpretation of the model as combining components to form larger elements, there are connections with two important literatures. Defining elements as strings of attributes is reminiscent of Lancaster's model of consumer demand in which products are bundles of attributes, and the actual value consumers place on a product is based on its underlying attribute bundle (Lancaster 1966). Lancaster and the extensive literature that follows does not focus on the ordering of a set of attributes in defining the value of a product. More fundamentally, this literature does not consider the creative process through which attributes are bundled together to create new products. The approach in this paper can be viewed as building on the product attributes literature, focusing on the creative process through which new products or more general cultural elements are created.

---

[1]In a separate work, Arthur Campbell and I (Campbell and Feinstein 2013) have decoded the set of concepts in this book and other microeconomics textbooks, uncovering more than 1200 concepts in total; however, we have not taken the step of organizing the material as a formal knowledge structure.

[2]The number of letters or attributes is ultimately not critical, since, as discussed in the theory of complexity (Li and Vitanyi 1997), alphabets can be translated into one another. From this viewpoint, the most basic alphabet is the 0/1 binary alphabet and any list of attributes can be redescribed as strings of zeroes and ones. However, a particular alphabet typically has a natural interpretation. Furthermore, the rules for which strings can be combined described in "Success Rate and Value Distributions," which are transparent for the given alphabet, would be more complicated if translated to hold for another alphabet.

[3]In a more general model, sets of letters are linked to attributes, so that for example a string of letters can produce an emergent attribute not present in any of the letters individually.

A second important interpretation is provided by viewing the model from the perspective of the knowledge representation paradigm. Knowledge representation is the study and development of frameworks for classifying knowledge and describing elements and how they relate to one another. An element in a knowledge representation has a set of components and attributes associated with it that define the element. The knowledge representation for a field will generally include basic elements, attributes, and more complex compound elements built up out of simpler elements, ultimately leading to the description of quite complex elements or systems. For example, in electronics, integrated circuits are often described in terms of components, the properties of those components, and circuits that define how a set of components are interconnected, together with properties of the circuit as a whole (see Dorf and Svoboda 2006 for textbook treatment). This is a more complex structure and representation than the string formulation in this paper, especially in being inherently two dimensional and with a sharper distinction between components and attributes; but it is consistent with the general approach adopted here. It is also interesting to note that the field of integrated circuits has developed enormously, and complexity has increased dramatically, including the development of very large-scale integrated circuits (VLSI) (Mead and Conway 1980); this fits the approach in this paper in which strings get longer as the field develops, representing more complex elements. Other examples of application include visual art, in which often a two-dimensional representation of form and color is relevant, architecture, fields of design, chemical engineering, and musical composition. While the model in this paper is quite simple relative to these examples, it provides an initial bridge between economic and decision-based models of learning, the principle of creativity as generated via novel conceptual combinations, and knowledge representation.

Initially, the field is comprised of N_0 elements each of length two that have been created previously. In general, we specify a simple ring for the initial condition. For example for $N = 4$, we set $N_0 = 4$: $a_1 - a_2$, $a_2 - a_3$, $a_3 - a_4$, and $a_4 - a_1$. For $N = 3$ we also set $N_0 = 4$ with initial elements $a_1 - a_2$, $a_2 - a_3$, $a_3 - a_1$ and $a_1 - a_2 - a_3$, adding one additional element to the ring to provide more options in the first period. We do not consider how the initial strings have been created but take them as given.

Creating New Elements

The model of creativity in this paper is based on the definition of creativity as *connecting, combining or relating two or more elements that have not previously been connected, combined or related*. Specifically, individuals create new elements in the field by combining two previously created strings that have not previously been combined.

The definition of creativity as novel conceptual combinations is at the core of many definitions of creativity. In the psychology literature on creativity, it is the essential idea in the principle of spreading mental activation generating novel creative ideas by linking elements that are not typically associated in ordinary, narrower trains of thought. The principle of spreading activation generating novel combinations is central to Mednick's well-known remote associations (Mednick 1962) and de Bono's principle of lateral thinking (de Bono 1970), as well as Koestler's notion

of bisociation (Koestler 1964). Spreading activation leading to novel combinations has also received neuropsychological support. Kenett et al. (2014) show that individuals who are rated as more creative generate word associations that produce wider lateral neural networks as opposed to individuals rated less creative who generate narrower more rigid networks. Lindell (2011) reviews the literature on network activation and argues the evidence shows that more creative thinking is associated both with the right hemisphere—typically associated with wider, less analytic processing, and with networks that cross the two hemispheres (right and left) thus are quite wide. More broadly, viewed as connections among elements in a field, creativity as a novel connection of two or more elements is the basis of many if not most innovations. A few examples are: hybrid vehicles, transistor radios (SONY), microwave ovens, and Darwin's theory of evolution by natural selection (link with Malthus) (see Feinstein 2006, Chapters 9, 10, 11, and 15 for many additional examples). Creativity based on novel combinations has in fact a wider application than might at first appear. For example, it incorporates connecting a novel solution with a problem (Fauconnier and Turner 1998; Poincaré 1908, 1952), metaphor (Gentner 1983), connecting a theoretical framework with a novel application, employing a conceptual schema to make sense of an experience or sensory data, and connecting a concept with its physical manifestation via an explicit technology.

The idea of creativity as produced by novel combinations of elements has been introduced in the economics literature by Weitzman (1998) and Feinstein (2006). However, it has not found its way widely into the economics literature thus far. Most commonly in economic models, a formal model of the creative process is not specified. Often, the focus is on how many resources are devoted to innovation with the basic assumption that more resources increases the likelihood of innovation, but without specifying an explicit underlying model of the creative process. An important literature has emerged that describes the creative process as search drawing from a distribution of possibilities (Evenson and Kislev 1976; Kortum 1997). While this literature does provide a more explicit model of how innovations are generated, it is not rooted directly in the creative process as making novel connections and combinations, and does not incorporate a knowledge representation structure of the field in which the innovations are sought. An important goal I have in this paper is to demonstrate the value of bringing a more structural model of the creative process to economics and allied subjects, in particular for predicting patterns of creativity and innovation in fields.

Although the basic definition of creativity as novel combinations is widely accepted, it is important to recognize that creativity can be viewed from a variety of perspectives, which are different from the approach adopted here, though complementary in most cases. The personality approach emphasizes individual differences in creativity; this includes creativity tests (see Torrance and Ball 1984 for the well-known Torrance tests of creativity) and personality structures or attributes that are conducive to creativity, such as openness to new experiences and ideas (Barron 1969). The role of motivation in creativity is also an important focus of study. There is a significant literature on the importance of intrinsic motivation and environments that are conducive to it for creativity generation, including a large literature studying the interplay of intrinsic and extrinsic motivation and the possibility that extrinsic motivators

like incentives may crowd out intrinsic motivation (Amabile 1996). I do not explicitly model personality, e.g., individual differences, or motivation factors in this paper. I assume that all individuals are ex ante identical in terms of creative potential—what is key is the state of the field when an individual enters and the intuitive signals an individual happens to receive, which are not tied to any prespecified personality traits, though they might be in extensions of the model. I further assume individuals earn a value from an element they successfully create and do not define further the source of value; thus, value may be based on both extrinsic and intrinsic factors. Extensions of the basic model can consider separately intrinsic and extrinsic factors that may influence values and thus creative development. For example, in an extension I discuss in the conclusion I have explored the role of royalties in the model and how they impact field development. Creativity can also be viewed from a more sociological perspective, as for example in Florida's well known work on the rise of the creative class (Florida 2002). Again, I do not take a sociological perspective—rather my approach is based on an individual perspective rooted in economics. I share with the sociology approach a focus on the environment as important in driving creativity. However, I frame the environment in terms of knowledge representation, which maps naturally into the individual rational learning process that drives model dynamics, rather than sociologically in terms of for example class identity or power that might restrict creative opportunities. Finally, creativity can be viewed from a systems perspective as the interaction of person, culture, and environment (Csikszentmihalyi 1988). This approach is relevant for the model in this paper, in which individuals generate creative combinations through direct interaction and learning in their cultural, field environment.

Notwithstanding the fact that the the model in this paper is based on a specific model of creativity generation, it generates a rich set of interesting empirical predictions as detailed in "Simulation Results."

Conceptual Bridges and the Creation of New Strings

Not all combinations are feasible. In this paper, I specify rules for which combinations are valid based on the cognitive science principle of conceptual blending (Fauconnier and Turner 1998). According to conceptual blending, conceptually valid combinations are made by connecting two concepts that share a bridge or overlap that enables their distinct elements to be linked. Thus, the combination of two technologies requires that they are able to be fit together via a bridge element, such as matching states or a viable physical linkage. As Fauconnier and Turner discuss at length, a valid solution to a problem requires that some conceptual frame be overlapping between the problem and the proposed solution, so that the solution fits the problem; likewise in metaphor, the "source" domain must naturally connect with the problem or "target" domain (Gentner 1983). A statistical model "fits" with a dataset if the data satisfies the conditions for correct application of the model: for example, if the model is a binary choice model then the data must also have the property that its dependent variable has two outcomes. The principle also applies in fields of design and technology, for example in electronics new circuits are created by linking existing circuits, and to be linked the components must fit together at the

Fig. 1 Combining strings, creating new elements

Simple RL Combination

$$a_1\text{-}a_2\text{-}a_3 \quad + \quad a_3\text{-}a_4\text{-}a_5\text{-}a_6$$

$$\|$$

$$a_1\text{-}a_2\text{-}a_3\text{-}a_4\text{-}a_5\text{-}a_6$$

bridge connection. More broadly, a creative association of any kind is possible only when the two elements being linked share some kind of point of connection, which could be experiential (such as overlapping in space and time) or conceptual (sharing terms or variables in common).

I apply this principle with the condition that *two strings can be combined into a new larger string when they share a letter or substring of letters in common, each having the letter or substring on an end; they can then be combined via their overlapping ends.* Figure 1 illustrates this combining process. Since each string has two ends, there are four possible ways to combine two strings, which I denote LL (the left edge of the first string connects with the left edge of the second), LR, RL, and RR.[4] Strings $a_1 - a_2 - a_3$ and $a_3 - a_4 - a_5 - a_6$ can be combined via the right edge of string 1 overlapping with left edge of string 2, a RL connection, to form the new string $a_1 - a_2 - a_3 - a_4 - a_5 - a_6$; this is shown in Fig. 1. Note that only a single a_3 is included in the final string as this element overlaps between the two strings. Strings $a_1 - a_2 - a_3$ and $a_4 - a_5$ cannot be combined. Strings $a_1 - a_2 - a_3$ and $a_1 - a_3$ can be combined in two ways: as an LL connection producing new string $a_3 - a_2 - a_1 - a_3$, and as a RR connection producing new string $a_1 - a_2 - a_3 - a_1$. Strings $a_1 - a_2 - a_3$ and $a_1 - a_2 - a_1$ can be combined in two ways, as a LL and a LR connection; however, both lead to the same final string: $a_3 - a_2 - a_1 - a_2 - a_1$. Overlaps of more than a single element are also allowed. Thus, strings $a_1 - a_2 - a_3$ and $a_2 - a_1 - a_4$ can be combined in the LL direction producing $a_4 - a_1 - a_2 - a_3$.[5]

New strings are always longer than either parent string. Thus, over time, longer more complex elements are created in the field. In this important sense, the field is non-stationary. This feature is consistent with the growth of many fields and human culture (as for example in Jones 2009), which clearly becomes more complex over time, notwithstanding that important simple ideas also continue to be added to the pool of creativity and innovations, here as new lateral strings, so that the field may grow in both depth and breadth.

The model of string elements can be generalized in several ways. One generalization allows attributes to be distinct from the building blocks (letters) used to construct elements. In this approach, there are two kinds of elements, letters and attributes, and

---

[4]There are two ways to write any given string as noted above. I assume that one way is chosen and fixed so that whether its left or right end is being used for a given pairing is well-determined. All subsequent strings are then checked against this string in both orientations. If a string matches in either direction, it is considered to be the same string.

[5]For a valid new combination, there must be at least one element from each parent that is not contained in the other parent, so that each parent contributes at least one new letter to the new string.

each string of letters has a set of attributes associated with it. The main advantage is greater flexibility, specifically in how attributes change as strings are combined to form new elements. In particular, some attributes may drop out and new "emergent" attributes may be associated with a new element. For example, a chemical compound created out of two component chemicals may possess new attributes neither parent possesses.

A second generalization is to specify elements using functions, rather than strings. Rules can then be specified that specify which building block elements can be linked to create new elements, more flexibly than the simple edge conceptual overlap approach I use. This is the approach used in many areas of knowledge representation. For example, in natural language representation, rules of grammar are defined as functions that specify which kinds of elements can be combined to construct valid new statements, such as a noun with a verb with a direct object; see Helbig (2006).[6]

The Creation of Subbundles

An important feature of the model, linked to the intuitive signals that guide creative development, is the creation of subbundles created jointly with a main string. For example, linking $a_1 - a_2$ with $a_2 - a_3$ to create string $a_1 - a_2 - a_3$ may also create as a byproduct the subbundle consisting of a_1 and a_3. These two letters are not adjacent in the final string, but they are implicitly connected since they are embedded in the larger string and linked via the bridging element a_2. Therefore, if they have value as a joint pair that value can be realized via the creation of the parent string. In fact, of the four attribute bundles created by forming $a_1 - a_2 - a_3$, only two are novel: the full string and a_1 with a_3 (the other two are the parent strings, $a_1 - a_2$ and $a_2 - a_3$). Thus, this subbundle may be quite important for the total value created by the new string. The perspective on the creative process in this model is that when a new element/product is created, the main source of value created is not necessarily the full product (though it may be) but rather can be a subbundle of attributes that have never been jointly produced before. I do not impose a specific assumption about which is more important, but rather define the total value realized when a new string is created to be the sum of the value of the main string plus the values of all co-created subbundles. Since these values are all different and stochastic, any one of them can be the predominant value created. I impose the condition that the subbundles are not realized unless the main string is viable.

The creation of a given subbundle of attributes or elements may be the principle aim in creating a larger parent element. Consider for example a new consumer product. This product may incorporate many elements or attributes, but many of these may primarily serve as bridges or connectors that make the overall product feasible, but are not the main source of value. Rather, a certain specific subbundle of attributes may be the key value creator, a view that is consistent with the Lancaster model.

---

[6]In general, for two building block elements g_1 and g_2, a function $f(g_1, g_2)$ specifies whether g_1 and g_2 can be combined, and if so also specifies the set of attributes associated with the new element. This process is recursive: the new element itself can become a building block: $f(f(g_1, g_2), f(g_3, g_4))$. The model of this paper fits with this approach, as one simple case, within a large family of possibilities.

For example, a new car design will have many elements, but there may well be a few key features bundled together that create the underlying consumer value. Likewise, a chemical process that fuses $a_1 - a_2$ with $a_2 - a_3$ to create the new material $a_1 - a_2 - a_3$ may use a_2 as a bridge but may have as its primary aim to join a_1 and a_3, creating a new material that combines these two elements for the first time. A new drug compound may be viewed from this viewpoint, with two or more therapeutic agents bound together in a deliverable form. Innovation from this viewpoint is a combination of (i) identifying valuable bundles of attributes, and then (ii) exploring or experimenting to find a larger template of elements that can be created and that embeds the high value subbundle.

As strings get longer, there are more possible combinations of subbundles. It seems unrealistic that a very large number of new subbundles will be realized. Thus for example in a long string, it may be difficult for the value contained in two letters that are placed far apart in the string to be realized since these elements are not closely linked. In order to keep the number of subbundles stable (stationary), I impose the condition that the only feasible new subbundles that can be co-created are those based on combinations involving parents of the main string and their parents (grandparents of the newly created string).

Figure 2 illustrates the creation of subbundles. In the figure, parent strings $a_1 - a_2 - a_3$ and $a_3 - a_4 - a_5 - a_6$ are combined in a RL configuration. The first parent has been produced from what are now grandparents $a_1 - a_2$ and $a_2 - a_3$, and the second from grandparents $a_3 - a_4 - a_5$ and $a_5 - a_6$. In total, there are eight potential subbundles: each parent with the opposite parent's grandparents (four), and each grandparent with the opposite parent's grandparents (four). One example is the combination of the two circled grandparents: $a_1 - a_2$ with $a_5 - a_6$. In fact, the actual number of potential new subbundles may be lower for several reasons. If a subbundle is an exact duplicate of a preexisting field element or the main string it is being created with—meaning it is the same string with the same parents, created the same way—then it is not co-created since it cannot be created twice. If two subbundles are duplicates, then assuming the string defined by the two subbundles is viable just one copy is created. Lastly, if two subbundles duplicate one another but are created differently (different parents), then assuming the string they define is viable, both subbundles are created but the value associated with the string they define is added to the total created value just once.[7]

An important assumption is that subbundles can be created without a conceptual overlap. Thus in Fig. 2, combining $a_1 - a_2$ and $a_5 - a_6$ creates the subbundle $a_1 - a_2 - a_5 - a_6$. Note that some combinations will have an overlap if they straddle the overlapping portion of the two parents, for example $a_2 - a_3$ and $a_3 - a_4 - a_5 - a_6$ in the figure. Subbundles that do not share a conceptual overlap are not viewed as new strings that can be used to build further strings, even though they do create value. Subbundles that do share a conceptual overlap are treated as new strings in their own right and can be used to build further strings.

---

[7]Additional rules are (i) that two subbundles cannot be combined if one is a subset of the other; and (ii) if the concatenation already exists as an element in the field, it is not re-created. Note that each subbundle has a fixed orientation in the parent string. I impose the rule that new subbundles can be created only with the orientation they occupy in the parent string, which is why there are just eight candidates.

The importance of allowing subbundles to have realized values is that individuals' intuitions guiding their creative development are often (and, specifically, in the model in this paper) about these smaller combinations that they then seek to find ways to realize, see below.

Success Rate and Value Distributions

Not all potential new elements (that is, valid based on conceptual overlap) are viable. The fact that there is a conceptual bridge to connect two elements makes it possible that a new element can be produced, but does not ensure that this will be possible. For example, a pharmaceutical company may find a chemical bridge that in principle enables two molecules to be linked to produce a new drug, but the attempt may fail and the new drug cannot be synthesized. The probability of success of new combinations is a parameter denoted P_X, $0 \leq P_X \leq 1$. In some fields, we expect the rate of success to be high, for example in more theoretical fields in which the logic of the combination may be sufficient to guarantee, at least in many cases, that the new element is viable. But in most fields, including empirical and experimental fields, P_X is likely to be closer to zero. When a new element turns out not to be viable, its value is zero, and no new subbundles it contains are created. For new elements that are viable, the value of the new element is drawn from a distribution. I specify two distributions, low or ordinary and high. The probability an element has its value drawn from the high distribution is denoted P_H; the expected value for the high distribution is v_H times the value for the low distribution; the value of which is set from other constraints specified below. The same distributions apply to subbundles. The event that the subbundle is viable is independent of the event that the parent is viable, and if it is viable, its value and whether its value is drawn from the high distribution are independent of the value and high/low draw for the parent. In addition, the viability, values, and high/low draws for any pair of subbundles are independent.

I specify the value distribution in a manner that is consistent with the empirical literature on valuation of innovations and creative output, especially patent citations and revenues and scholarly citations. It is widely accepted that the distribution of values associated with innovations is skewed with a long right tail, with a relatively small percentage of innovations generating high value. Silverberg and Verspagen (2007) in a careful analysis of several different datasets find that the main body of the value distribution in a variety of applications is well fit by a log-normal distribution but that the tail of the distribution is better fit by a Pareto distribution.[8] To capture this empirical regularity, I specify a distribution with two parts that are spliced together to create a single value distribution. The first, lower part of the distribution applies for ordinary creative contributions and the second upper tail applies for contributions that have unusually high value. I specify the distribution of values for the main lower part of the distribution to be log-normal consistent with the findings of Silverberg and Verspagen, specifically a truncated lognormal distribution with truncation point X_m. The mean μ and variance σ of the distribution govern how skewed the distribution is

---

[8] See also Scherer and Harhoff (2000), Gambardella et al. (2008), Jaffe and Trajtenberg (2002).

and together with X_m determine the mean and standard deviation. For contributions with unusually high value, I specify that the value is drawn from a Pareto distribution with cut-off point X_m and parameter α. Parameters are chosen to create a well-formed distribution consistent with other model parameters. In particular, μ, σ, X_m, and α are chosen such that the overall cumulative probability is one, the cumulative probability associated with the Pareto portion is P_H, cumulative probability associated with the lognormal portion is equal to $1 - P_H$, the expected value for the Pareto upper tail is v_H times the expected value for the truncated lognormal, and the density function of the truncated lognormal at the point of truncation equals the density function for the Pareto at X_m, so that the two densities splice together in a continuous manner. Figure A1 in the Appendix depicts the resulting value distribution for the base parameter values shown in Table A1.

Creative Development

Individuals enter the field in sequence. Each individual who enters the field engages in a process of creative development consisting of four steps. First, he chooses a seed learning set from the existing elements in the field. The seed must consist of more than one element in order to generate intuitive signals as specified below; in the simulations, it consists of two elements. Second, he gains intuitive signals about creative opportunities in the field, specifically bundles of attributes. Third, guided by his signals, he chooses an additional set of elements to learn; in the simulations, this second learning set also consists of two elements. Fourth, he chooses a new element to attempt to make as his creative project from the set of potential new elements he can make given his full learning set. This potential new element must be based on combining two elements he has learned that share a conceptual overlap. The outcome of his choice is then realized. If the element is not viable, there is no value created and no new elements added to the field. If the element is viable, it is added to the field and the individual realizes its value. In addition, if the element is viable, outcomes are realized for any subbundles co-created with the main new element: each subbundle is revealed as viable or not, and if viable it is added to the field and the individual accrues its value.[9] Individuals make choices with the objective of maximizing the expected value they will earn.

Limited learning capacity is an important constraint in the model. As a field grows, it contains a large number of elements, and no individual can learn everything, especially not with the degree of understanding required to build creatively with an element. The choice of what to learn is thus critical. Limits to learning can arise and be imposed through various mechanisms. One approach is to specify a cost for each element learned, possibly linked to its complexity, in this model measured

---

[9]If the string has been made before and is thus already in the field no new value is created, even if the individual makes it a new way, e.g., via a different pair of parent strings. Though it is unlikely, it is possible for an individual to choose to try to make a string that has already been made, if he believes that the subbundles that would be co-created are sufficiently valuable.

by length; individuals then optimize subject to a learning budget. Rather than complicate the model to that degree, I impose the constraint that an individual selects and learns a fixed number of elements in each learning cycle. Since the length of selected elements may increase, this means that learning becomes more efficient as the field matures, in the sense that larger knowledge chunks (longer strings) are learned.[10]

A second related issue is which components an individual learns when she selects an element. I assume that an individual learns not only the element itself but also the two parents that were combined to form the element, but no additional elements beyond this. This assumption may not seem necessary. It might seem simpler to assume that an individual learns every component substring when she chooses to learn an element. However, as elements in the field become longer, they contain more and more substrings/subbundles—each parent has grandparents; these grandparents in turn have parents, and so on. Thus, the assumption that all substrings/subbundles are able to be learned would imply that as elements become longer an individual could learn a very large number of elements from selecting just a single main element, thus effectively circumventing the learning capacity constraint. This is one of several issues in which the non-stationary nature of a field has implications that model assumptions must address. The assumption I make preserves a degree of stationarity, in that the number of elements learned remains at three regardless of how long elements become. However, the length of these learned elements may increase over time, reflecting greater efficiency in the way in which knowledge is packaged into blocks. Aside from learning capacity constraints, the ability to deconstruct a given string and learn its component elements is undoubtedly limited. While in some cases, like a movie or simple toy, these components may be accessible, in many cases, such as composite materials, engineered products, or food, they may not be directly accessible or may have been transformed in such a way that they cannot be extracted (reverse engineered) and learned. Parents would seem to be the most accessible components since they have been used to construct the main element.

Two final comments. One, learning occurs in stages, not all at once, so that as an individual learns and gains intuition from what he has learned, this influences his subsequent learning choices. In the model in this paper, this is modeled in the simplest possible way, as a two-step learning process. Two, individuals must make commitments about which project to pursue since projects are costly and individuals cannot pursue all projects they may imagine. Here, I make the simple assumption that an individual can pursue only one project. It is straightforward to generalize the model to allow an individual to pursue more than one project and put into the field the best one; this has implications for creative development since when individuals can pursue more projects, they are more likely to pursue projects with high tail values but lower overall probability of success.[11]

---

[10]The model can be interpreted as the case in which the economies of learning for each element are large so that the full string is learned for a fixed cost. Even if a more explicit learning-cost model is employed, the cost of an element need not be linear in its length.

[11]The restriction to focusing on just one project incorporates in a simple way the crucial point emphasized by Weitzman that it is costly to pursue the development of ideas, so that even though many potential

Intuitive Signals

Based on their seed elements, individuals gain signals about the value of potential new bundles of attributes. Specifically, signals are associated with the bundles formed by concatenating an element from one seed element with an element from the other. There are three learned elements associated with each seed—the seed itself and its two parents. There are thus nine possible pairings (there can be fewer if there is duplication). Each such pair b_{1i} and b_{2j} can be concatenated in four different ways—LL, LR, RL, and RR. Hence there are as many as 36 potential string concatenations or subbundles that have signals associated with them for a given seed set. Importantly, these subbundles are not strings formed by conceptual overlap. Rather, they are bundles of elements that can generate values as subbundles of larger strings that are created.[12]

Each subbundle that has associated signals generates two signals. One signal provides information about the likelihood that this subbundle is viable. The other provides information about the likelihood that the value associated with the bundle, given it is viable, is drawn from the high distribution.

The fact that signals are associated with subbundles is an important feature of the model of creative development in this paper. Due to this feature, the main way signals enter into creative development is through being associated with bundles of elements that are created as subbundles embedded in larger strings. The logic motivating this feature is that intuitions guiding creative development are in most cases not about complete products, with every detail worked out, but about possibilities that are not fully formed, but only partially imagined. Referring to the Lancaster model of attributes, one may have an intuition that a certain bundle of attributes, if combined will have value. One then seeks out elements containing these attributes that fit together to create a viable full product. The product undoubtedly contains many additional elements, various "connectors" used to embed the attribute bundle one believes has value. But much of the value of the final product in fact derives from the specific combination of attributes one envisioned. Creative development, from this point-of-view, is about intuitions about simpler combinations, then searching for building blocks within which, when they are fit together, these smaller bundles are embedded.[13] Creative development starts from broader imagined possibilities and involves the search to develop this broader intuition more completely, ending in a workable final creative product.

---

combinations are available, with the number increasing dramatically as the field grows, only a relatively small number and smaller percentage over time can actually be pursued.

[12]There are two kinds of concatenated strings that do not have signals associated with them. An ordered bundle that has already been produced as a string with the same elements in the same order does not have any signals associated with it, since its value is already known. And signals are not associated with strings for which the two parents share a conceptual overlap on their facing ends as these two elements can potentially be produced as a new main element. It is not difficult to extend the model to allow signals to be generated for such conceptual overlap strings.

[13]It is possible for a string that has been associated with a signal via concatenation to be produced via conceptual overlap through combining two different parents. In this case, the signal has provided information about a final main element that is produced, thus the model allows for this possibility.
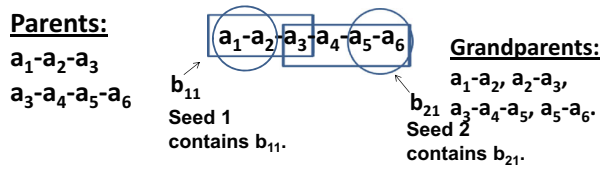
Parents:
a_1-a_2-a_3
a_3-a_4-a_5-a_6

$(a_1$-a_2-a_3-a_4-a_5-$a_6)$

b_{11}
Seed 1
contains b_{11}.

Grandparents:
a_1-a_2, a_2-a_3,
b_{21} a_3-a_4-a_5, a_5-a_6.
Seed 2
contains b_{21}.

Fig. 2 Subbundles and signal-generating blocks

Imagine if instead of the model I propose here intuitions provided information about the value of a complete new string element and the exact components needed to produce it via conceptual overlap. Creative development would then be focused solely on searching for these components enabling the element to be constructed exactly as envisioned. The fact is that most intuitions in the course of creative development are not this sharp, but leave room for different approaches for how an intuitively valuable combination will be realized. An engineer does not imagine every detail of a new product when he first conceives of its possibility. Rather he imagines certain critical features that will be incorporated. Though it might seem simpler as a modeling strategy, it is simply not correct to reduce creative work to a by rote search for a fixed set of elements aiming at a fixed final product. One must leave room for the ongoing process of exploration and adaptation as one moves from initial conception, often just a partial vision, to a complete final product.

A related implication of the model is that the intuitions guiding creative development are often at a more abstract conceptual level. In the knowledge representation framework, a short string or smaller bundle refers to a more basic concept, whereas a longer string embedding this string has added more elements refining the basic concept. For example, a short string might define a simple circuit or chemical compound, and a longer string that incorporates this shorter string defines a more complex circuit or compound. Thus, an intuition about a relatively short string is about a relatively broad conceptual combination. An example is an intuition that it will be fruitful to link a theoretical or statistical framework with a new area of application. A researcher may believe that this will be fruitful. His intuition, however, is not so fully developed that he knows exactly what dataset he will use and what the exact theoretical, statistical model will be within in the broader family that fits with the dataset he ends up using. He then searches for a specific model and dataset that he is able to fit together and that incorporate the broader link he envisions. The principle that creative development is guided by broader conceptual interests and intuitions is developed and demonstrated with many examples in Feinstein (2006). Among many examples discussed there are: Alexander Calder's creative interest in the universe (the solar system) as the basis for art; Matisse's conception of combining vivid, strongly contrasting colors cutting across boundaries of form (see also Spurling 1998 and Matisse 1990); and John Maynard Keynes' initial thoughts about the relationship among expectations, investment, and economic fluctuations (see also Skidelsky 1983; Keynes 1977, 1981).

Figure 2 illustrates how a subbundle associated with signals is embedded in a larger string. In the figure, seed element 1 contains block b_{11} and seed element 2

contains block b_{21}, each of which refers to a grandparent. The subbundle formed by concatenating b_{11} with b_{21}, string $a_1 - a_2 - a_5 - a_6$, is a possible signal-generating subbundle since its two blocks come from different seed elements. If the main string shown is attempted as a project and turns out to be viable, then this subbundle will also be formed if it is viable. Thus, if an individual receives a positive signal that this subbundle is likely to be viable, that may encourage him to try to make the main string, since if it turns out to be viable the subbundle is likely to be co-created, adding additional value. Indeed, it is here that intuition guides creative development: the value that may be (and is likely with a positive signal) to be created by the subbundle may be the main driving factor behind the decision to attempt to make the full string.

The individual gains signals about m concatenations from his seed set. In the simulations, m is set to 2. The concatenation subbundles for which he gains signals are chosen at random from among the set of valid signal concatenations. For each such subbundle, he receives two signals, a signal about the viability of the subbundle, called the X signal, and a signal about the likelihood that the value of the subbundle if it is viable is drawn from the high distribution, the H signal. The X signal is either 1 (viability likely) or 0 (viability unlikely). Signals are not perfect: The false positive rate for an X signal is s_0^X, and the true positive rate is s_1^X. Likewise, the H signal is either 1 (string value more likely to be drawn from the high distribution) or 0 (relatively unlikely). The false positive rate for the H signal is s_0^H and the true positive rate is s_1^H.

For a given concatenation string, the X and H signals are independent (the H signal is, however, relevant only if the string is viable). Furthermore, for any two distinct subbundle concatenations, their respective signals are independent. If a given subbundle can be made via concatenation two different ways, then in any case in which signals are generated for each of two ways of making it the pair of X signals generated are conditionally independent conditional on whether or not the subbundle is viable, and likewise the pair of H signals are conditionally independent conditional on whether or not the subbundle value is drawn from the high distribution. I provide formulas for learning rules covering these cases in the Appendix.

Optimal Strategy

The optimal strategy for an individual must specify (i) which seed set he selects; (ii) for each possible signal pair that may be generated based on the seeds and each signal draw for that pair of signals: (a) the additional elements he chooses for his full learning set; and (b) the new element he chooses to try to make as his creative project.

The first decision is the selection of seed elements. Assuming there are N_t elements in the field at the beginning of period t and the individual selects two, there are $N_t(N_t - 1)/2$ possible seeds. For each possible seed, the individual computes the expected value associated with the optimal strategy if he chooses this seed; he then selects the seed with highest expected value.

Given a seed choice, the individual will receive signals from m members of the signal concatenation set, drawn at random; I set $m = 2$ for the remainder of the discussion to make the formulas simpler. If there are m_{ts} valid signal concatenations for seed s, there are $m_{st}(m_{st} - 1)/2$ possible signal combinations, with each combination

equally likely. The individual computes the highest expected value he can gain for each of these possible combinations. The steps involves in this calculation are outlined below. He then averages these values to compute the expected value associated with this seed.

Each signal concatenation has four possible signal outcomes: $X = 1$ and $H = 1$; $X = 1$ and $H = 0$; $X = 0$ and $H = 1$; and $X = 0$ and $H = 0$. Signals are generated in pairs for $m = 2$ hence there are 16 different possible combinations for a given signal draw. Denote the prior probability that string s is viable by $P_X(s, t)$. I use the terminology "string" and note these formulas apply to both main strings and sub-bundles or concatenated strings. Likewise, denote the prior probability that the string has its value drawn from the high distribution by $P_H(s, t)$. Note that string probabilities are updated after each period based on what is observed about the behavior of the individual who worked in the field; update formulas are given in "The Development of the Field." Define $psig_X(s, t)$ to be the probability that the signal $X = 1$ is generated for concatenated string s and $psig_H(s, t)$ to be the probability the signal $H = 1$ is generated. These probabilities are:

$$psig_X(s, t) = s_1^X * P_X(s, t) + s_0^X * (1.0 - P_X(s, t)) \tag{1}$$
$$psig_H(s, t) = s_1^H * P_H(s, t) + s_0^H * (1.0 - P_H(s, t))$$

The calculations of the probabilities associated with each set of signals is now straightforward. When the two concatenated strings for which signals are generated are different, the signals for the first string are independent of the signals generated for the second, making the calculation an easy set of multiplications. As an example, for a pair of concatenation signal generators s_1 and s_2, the probability that the signals for the first string are $X = 1$ and $H = 1$ and the signals for the second pair are also $X = 1$ and $H = 1$ is:

$$psig_X(s_1, t) * psig_H(s_1, t) * psig_X(s_2, t) * psig_H(s_2, t) \tag{2}$$

The case in which the two strings s_1 and s_2 are the same string (made two different ways) involves somewhat more complex formulas provided in the Appendix.

After an individual receives his signals, he updates probabilities for the associated concatenation strings. Updating is done using standard Bayesian formulas. If the individual receives a signal $X = 1$ for the string, his revised probability that it is viable is:

$$P_X(s, t \mid X = 1) = \frac{s_1^X * P_X(s, t)}{psig_X(s, t)} \tag{3}$$

If the individual receives the signal $X = 0$ his revised probability that the string is viable is:

$$P_X(s, t \mid X = 0) = \frac{(1.0 - s_1^X) * P_X(s, t)}{(1.0 - psig_X(s, t))}.$$

Similarly:

$$P_H(s, t \mid H = 1) = \frac{s_1^H * P_H(s, t)}{psig_H(s, t)}$$

$$P_H(s, t \mid H = 0) = \frac{(1.0 - s_1^H) * P_H(s, t)}{(1.0 - psig_H(s, t))}.$$

When the two strings for which signals are generated are the same string (made two different ways), the formulas are more complex and are again provided in the Appendix.

In general, the most likely set of signal values is the one for which all signals are zero, due to the fact that for the parameter values used in the simulations the majority of strings are not viable and the likelihood of having a value drawn from the high distribution is low.[14] In this case, an individual's signals act as negative information, and may lead him not to choose a full learning set he would have chosen if he had not received any signals or not to try to make a new element that he would have tried to make if he had received no signals. Although this is the most common occurrence and therefore important for how the field develops, the more interesting cases are those in which an individual receives at least one signal that is a 1, in which case he may well be led to make choices such that he attempts to make a new element that has the string with which the signal is associated as a subbundle. This fits the commonsense view that individuals are guided towards elements that they believe have creative potential. In fact, much of the time a signal has no import, in that the subbundle it provides information about cannot be created given the current state of the field. I call such a signal a *clean signal*. As a related matter, an important benchmark of the model is what I call a *clean* run, in which individuals receive no signals to guide them, but make decisions working solely from public knowledge. It is interesting to compare how the field is projected to develop in this case versus when individuals do receive signals—I explore this comparison in the simulations.

An individual chooses two additional elements from the field to complete his learning set. Given that two elements from the field have been chosen for the seed, he chooses from among the $N_t - 2$ remaining elements. Given the choice of the full learning set, the individual determines all possible new elements she can make via conceptual overlap from among the elements in her learning set. There are up to 12 distinct elements in the learning set (there can be fewer if some elements are duplicates), and each pair can be combined in four different ways, thus there are up to $66 X 4 = 264$ potential new elements. Most of these combinations do not share conceptual overlap. Furthermore, among those that are valid new combinations some duplicate elements already in the field. Thus, the actual number of elements in the new element set is in general well below this, typically no more than a few dozen. The individual computes the total expected value associated with each potential new element, including the expected values associated with all potential subbundles, then chooses the element with highest expected value to attempt to make. If the element turns out to be viable, then it is created along with all viable subbundles, and the total value is realized and accrues to her.

The expected value associated with an element includes the expected value of the element itself as well as the expected values for each subbundle associated with the element that may be co-created with it. These expected value calculations are based

[14]This is true initially. However, as the field develops, some strings may have updated prior probabilities that sharply increase the probability that they are viable or the probability that they have a value drawn from the high distribution, overturning this general tendency.

on the probability assessments the individual makes that a given element is viable and the likelihood that its value is drawn from the high distribution. For a main element and any subbundles for which the individual has not received any intuitive signals, these probabilities are the prior assessments $P_X(s,t)$ and $P_H(s,t)$. For a subbundle for which she has received signals, the probabilities are based on the posterior probabilities above.[15] Assuming a string has probability $P_X(s,t)$ of being viable and probability $P_H(s,t)$ of having its value drawn from the high distribution (posterior probabilities after all signal updates), the expected value associated with the string is:

$$P_X(s,t) * \exp(\mu + 0.5 * \sigma^2) * (P_H(s,t) * v_H + (1.0 - P_H(s,t))$$

Given the calculation of the optimal new element to attempt to make and its expected value, the individual rolls back to compute the optimal full learning set given the signals he receives, and averages over all possible signal combinations to evaluate the expected value associated with a given seed. Finally, he rolls back to determine the optimal seed to choose.

The Development of the Field

One individual enters the field each period and lives a single period. The individual determines his optimal strategy as described in the preceding section. If the new element he attempts to make is viable, it is added to the field along with all associated new viable subbundles; the main element and any subbundles formed through conceptual overlap become new potential building blocks to create further elements.

Since each element is made through combination of two preexisting elements, the field has a structure resembling a lattice, specifically a semi-lattice, assuming a single unitary element that then divides into the N attributes.[16] In general, the field can grow without bounds.[17] It grows both in terms of depth or complexity, as new longer elements are created, as well as in breadth. Overall, the nature of growth is non-stationary in that new elements become longer with no bounds to how long they can become. It is possible to work out analytic formulas for how the field develops in simplified settings. One simplification is to assume that given the current state of the field, every feasible new combination that can be made, based on the rules for overlapping above, is attempted. In this case, when the viability probability is 1,

---

[15]If two subbundles duplicate one another, then if one has signals associated with it but not the other, the standard signal probability formulas apply; if both copies have signals associated with them, then the more complex formulas given in the Appendix apply.

[16]It is easy to see that the lattice is modular since every element has two parents. I do not explore the possible lattice properties of the field in this paper but this is a potentially fruitful line of investigation for future work.

[17]It is in theory possible to reach a position in which no new conceptual overlap combinations can be made, in which case the field growth ends. However, this is very unlikely especially once the field begins to grow beyond its initial conditions.

so every feasible new element is viable, the formulas are especially simple and are provided in the footnote to this sentence.[18]

The individual who enters the field in period t is assumed to be able to observe the history of the field and the choices made by previous individuals who worked in the field. I assume specifically that the individual observes the following for each prior individual: the seed and full learning set the individual chose and the new element he tried to make and whether it was successfully made or not. I assume the individual does *not* observe the intuitive signals prior individuals received—signals are treated as private information not shared through public information sources. Using the information he has access to the individual forms updated probabilities, for each possible new element he may make, including subbundles, of the probabilities the element is viable and has a value drawn from the high distribution, denoted $P_X(s, t)$ and $P_H(s, t)$ for string s (note that these are the probabilities prior to any signals the individual receives, which prompt further updates for certain subbundles). The updating can be worked out as recursive formulas, in that the update probabilities computed by the individual who entered the field in period $t - 1$ serve as prior probabilities for updating by the individual entering the field in period t. Thus, I give the formulas for a single period or round of updating.

The seed choice made by the individual who worked in the field in $t - 1$ could have been exactly predicted based on the $t - 2$ period choices and outcomes since individuals entering the field do not have any private information. However, the full learning set the individual chose in $t - 1$ and his selection of which new element to try to make in general depend on the signals he received. Furthermore, more than one set of signal combinations can lead the individual to choose the given full learning set he chose and new element he chose to try to make. Hence updating is done based on pooling all signal combinations that would have led the individual to make the choices he made. To simplify exposition, I denote this pool of signal combinations that is consistent with the observed choices of the individual in $t - 1$ by *pool*. Let P_{pool} be the total probability summed over all signal combinations that fall in the *pool*:

$$P_{pool} = \sum_{i \in pool} P(sig\ combo\ i)$$

where $P(sig\ combo\ i)$ is computed based on Eqs. 1 and 2 in "Optimal Strategy."

Updating proceeds in two steps. In the first step, probabilities are updated for all subbundles that have signals associated with them in the *pool*. For each such subbundle s, determine the set of combinations in the *pool* for which the signal received in regards s was a 1 (set 1), the set of combinations for which the signal received

---

was a 0 (set 2), and the set of combinations for which no signal was generated for s (set 3). Let q_1 be the weighted probability of the first set relative to the overall pool probability:

$$q_1 = \frac{\sum_{i \in pool} Ind(sig\ combo\ i\ has\ X = 1\ for\ string\ s) P(sig\ combo\ i)}{P_{pool}}$$

In this expression, the numerator uses an indicator function to identify which signal combinations fall in the set 1 and weights each such combination by its probability. Similar expressions hold for q_2 and q_3 for sets 2 and 3. The updated probability that s is viable is then computed as:

$$P_X(s, t \mid X = 1) * q_1 + P_X(s, t \mid X = 0) * q_2 + P_X(s, t) * (1.0 - q_1 - q_2)$$

The first two terms in this expression use the formulas for updated probabilities based on signals from Eq. 3 in "Optimal Strategy," and the last term uses the prior probability for s since this refers to signal combinations for which no signal was generated for string s. Comparable formulas hold for updates for the probability that the value associated with s is drawn from the high distribution. When two signals in a given signal combination refer to the same string, the formulas are more complex and are provided in the Appendix.

In the second updating step, probabilities are updated for the new element that the individual working in the field in $t-1$ tried to make and for all associated subbundles. When the new element was successfully made, its viability probability is updated to 1.0 and since its value is now known, the probability its value is drawn from the high distribution is no longer relevant and can be discarded. In this case, all subbundles are revealed as either viable or not—if a subbundle is viable its value is revealed and if not its value is set to zero. In either case, the probability its value is drawn from the high distribution is no longer relevant. When the new element was not successfully made, its viability probability is set to 0.0; its value is no longer relevant since it is not viable and the probability its value is drawn from the high distribution is also no longer relevant. In this case, nothing is learned about the subbundles associated with the new element—since the new element was not made, there was no opportunity for the subbundles to be co-created with it. Hence, the probabilities associated with the subbundles remain as they were after the first update step.

Simulation Results

I analyze the model through extensive computer simulations employing the analytic framework and expressions given in the preceding sections. The analysis is performed by a suite of FORTRAN programs, and has been executed on the Yale High Performance Computing System. The simulation protocol is described in the Appendix.

Parameter values have been chosen to focus on exploring the way intuitive signals guide create development and influence development of a field. Thus, the signals are of high-quality with low false positive and high true positive rates. For signals of viability, the false positive rate is set at .05 and the true positive rate at .9, and for
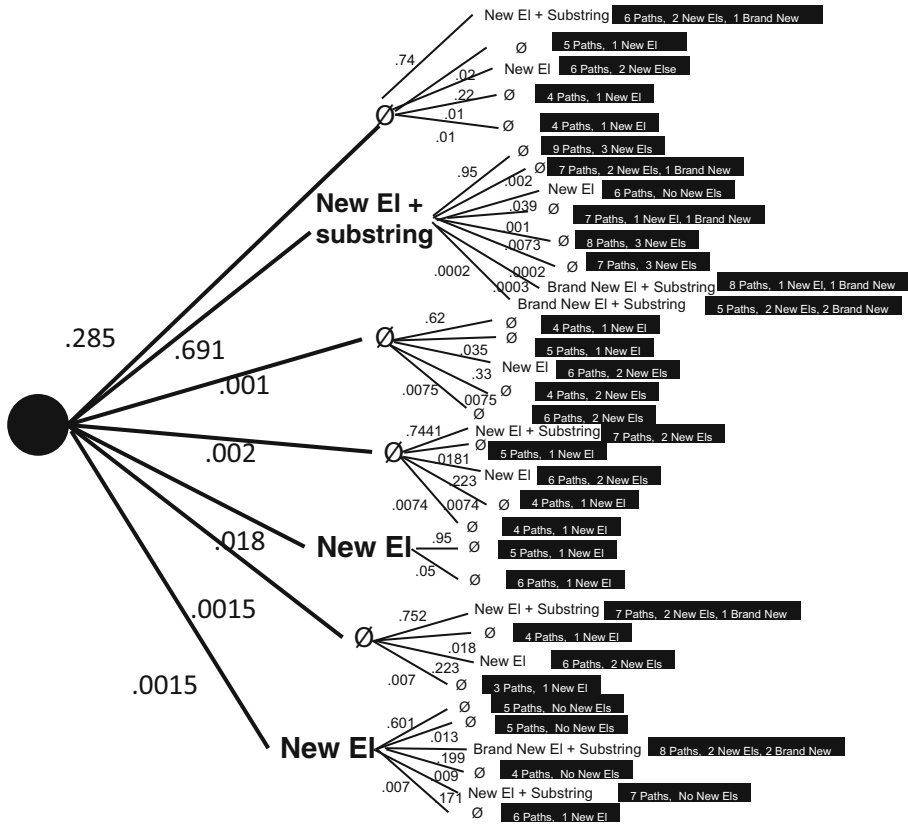
Fig. 3 Tree showing paths of development of the field

signals of a draw from the high distribution the false positive rate is .1 and the true positive rate is .9. Table A1 in the Appendix lists all parameter values used in the simulations.

I present detailed results here for one simulation run for $N = 3$, then discuss the pattern of results across all simulations. The initial state is shown in Fig. 4 on the top left diagram. Figure 3 depicts a tree showing paths of development of the field for this run. The development through the first two generations is shown in some detail; generation 3 is shown more schematically. For the root node, there is a single optimal seed. There are seven paths emanating from the root node, which differ in terms of which new element the individual tries to make; these different choices in turn are driven by different possible signals (recall that every feasible signal pair for the given seed is simulated, so every possible path of development is identified). For this run, based on the *masterlist*, three of the main elements turn out to be viable, paths 2, 5, and 7, and one of these spawns an additional new subbundle. Note that for the four nodes with no new elements, their histories are still different and thus the further development of the field may be different; most obviously which element was attempted in period 1 is different, and no individual along a path will attempt again an
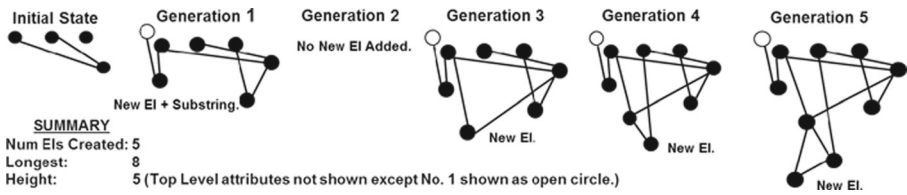
Fig. 4 Field structure development: example

element that has already been attempted and shown not to be viable. The probabilities are quite high for the first two paths, the first associated with what I have called *clean* run for which the signals are not directly relevant, the second associated with many of the other signal draws. The remaining five paths have far lower probabilities; each of these paths is associated with specific positive signals that are drawn with low probability. In fact, this is a typical pattern at most nodes for most simulations: 1 or 2 paths of relatively high probability, the remainder with quite low probability.

In generation 2, there are a total of 35 paths; 12 lead to the creation of one or more viable new elements, and 3 of these 12 generate an element that is brand new in the sense that it was not created along any path in generation 1—all of these are paths for which a new element had been created in period 1, opening up new possibilities in period 2. The most paths emerge from node 2, for which a new element was created in period 1, reflecting the greater richness and options in the field as elements are added. Overall, through just two generations, clear differences emerge in how the field is developing depending on intuitive signals the individuals working in the field receive and the choices they make based on these signals. In period 3, there are 199 paths. On average, there are more paths emanating from nodes for which more viable elements have been created in the preceding generations, with the greatest number from node 2. This is because when there are more elements in the field, there are more possible combinations that can be formed, hence more valuable signals, in turn triggering greater variety of choices of full learning sets and new elements to try to make. There are never more than 9 paths emanating from a given node, and in fact it is generally rare to have more than 12 paths emanating from a node. This is so even though the number of distinct signal combinations can be considerably higher; in many cases, subsets of signal pairs lead to the same choice of full learning set and new element to attempt and therefore are pooled for the *history* of the field as discussed in "The Development of the Field." I do not exhibit generations 4 and 5 on the tree due to the large number of paths. Through generation 5, the number of paths in total is more than 6000 for this simulation. The number of distinct field structures that is created is in general far smaller than the total number of paths because many paths lead to the same structure in terms of which elements have been created through period 5. For this simulation run, there are 131 distinct field structures through period 5.

Figure 4 depicts the development of the field for one illustrative path. In period 1, a new element is created and a subbundle is co-created. In period 2, the new element that is attempted is not viable. In each of periods 3, 4, and 5, a new element is added.
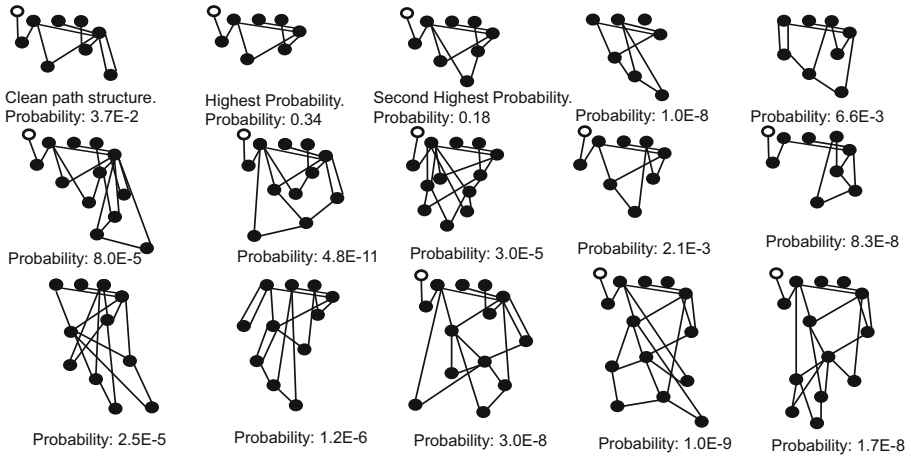
Clean path structure.
Probability: 3.7E-2

Highest Probability.
Probability: 0.34

Second Highest Probability.
Probability: 0.18

Probability: 1.0E-8

Probability: 6.6E-3

Probability: 8.0E-5

Probability: 4.8E-11

Probability: 3.0E-5

Probability: 2.1E-3

Probability: 8.3E-8

Probability: 2.5E-5

Probability: 1.2E-6

Probability: 3.0E-8

Probability: 1.0E-9

Probability: 1.7E-8

Fig. 5 Field structure: examples

In period 3, the new element is created using a pair of the initial field elements. In period 4, the new element combines an initial element with the element created in period 3, showing how individuals build on previous work in the field. Lastly, in period 5, the new elements created in periods 3 and 4 are combined. The final structure has five elements and a height of 5.[19] It uses all four of the initial field elements and contains as the most complex (longest) element a string of length 8.[20] The structure provides a basic empirical description of the field—it consolidates all paths that generate the structure, which is sensible for empirical description of field development.

Figure 5 depicts examples of 15 other field structures. For each structure, its cumulative probability summed over all paths that lead to it is listed. The first structure shown is associated with the path of *clean* signals every period.[21] The second and third structures show the structures that have the greatest cumulative probabilities, 0.34 and 0.18. These are structures where just a few elements that are attempted turn out to be viable—thus many different paths, with different histories of failed attempts

---

[19]Height is measured as the longest chain, beginning from an attribute; note that the attributes are not shown in the figure except for attribute 1, shown by an open circle, utilized in the creation of the subbundle.

[20]Multiple paths generate this same final structure, differing in the signals received by individuals working in the field and the order in which elements are attempted to be made, as well as and including different elements that are attempted but turn out not to be viable. Note thus that two paths that generate the same field structure nonetheless represent different *states* in terms of the probabilities of viability and of having a value drawn from the high distribution for strings that have not yet attempted, thus may lead to different choices in subsequent periods.

[21]This structure is also created along some non-clean signal paths, and the probability shown is the cumulative probability along all paths that lead to the structure. Note also that the bottom element is composed of initial element 4 combined with itself (in reverse order, so 1-2-3-2-1). That is possible because a new element added earlier has been made using initial element 4, thus two copies of element 4 are available. This is not unusual in the development of the field and satisfies the rules for constructing new elements laid out earlier.
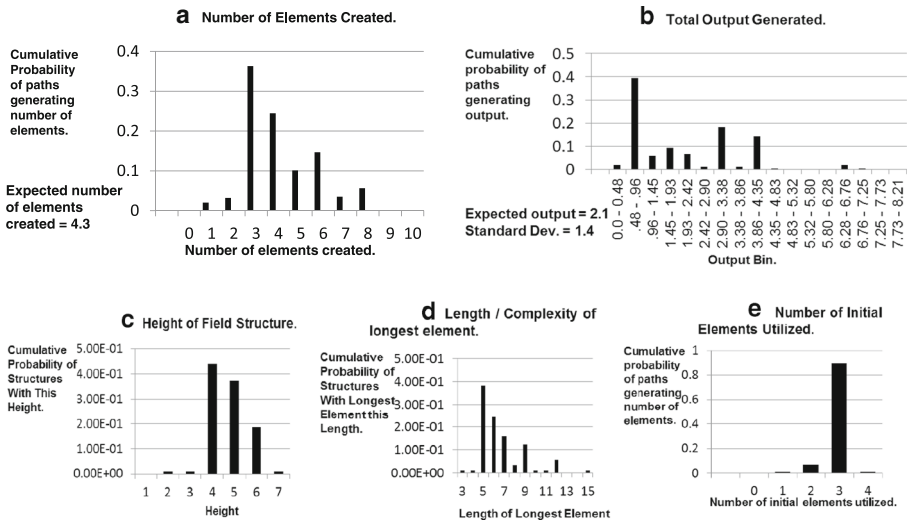
Fig. 6 Distributions

to make other elements, pool into these structures. The remaining 12 structures provide examples drawn from the 67 structures out of the total 131 that are uncovered in that they are not contained in any other structure in the set. These structures thus typically contain more elements, though I have selected examples that cover the full range of number of elements, from 3 to 10. Most of these structures are rare, generated along low probability paths typically tied to positive signals that lead to distinctive choices for which new element to attempt to make; however, many of these structures subsume structures with fewer elements that have substantially higher probability.

Figure 5 depicts how diverse the field's development can be, both in terms of specific elements and statistical measures such as height. This finding of significant diversity highlights the wide range of possible paths of development of the field, an important point of this paper. In some cases, one initial element is pivotal. For example, structure 5 (first structure, second row) shows a pattern of development in which element 4 plays a central role, being the parent for all but one of the subsequent elements created; similarly, in structure 7, element 1 plays a central role. Structures on the bottom row show more complex elements playing a central role, stretching the structure vertically. Every path is generated through optimizing behavior on the part of the individuals working in the field, governed by the specific starting node and intuitive signals these individuals receive. Thus, this diversity is not due to non-optimizing behavior, but rather to differing intuitive signals about valuable subbundles, which in turn guide the creative process. The diversity is maintained and indeed grows over time at least through period 5—the number of distinct field structures is more each period.

Figure 6 presents statistics about distributions for the field structures generated in this simulation, based on the cumulative probability of all paths that lead to a given outcome. Figure 6a shows the distribution of the number of new elements created.

The mean number of elements created is 4.3. The modal number of new elements is 3, associated with structures having a cumulative probability of .36. There is also substantial probabilities associated with 4, 5, and 6 elements being created, low but significant probabilities associated with 7 and 8 elements, and very small but nonzero probabilities for 9 and 10 elements being created—paths for which several subbundles have been co-created with main elements. Figure 6b shows the distribution of output. Mean output is 2.1, the standard deviation is 1.4, and maximum output is 8.0. The distribution is multi-modal due to the fact that high value elements are created on a few high probability paths. The highest probability bin is a relatively low output of between approximately 0.48 and .98 with an associated probability of 0.4. Most of the paths in this bin are paths generated based on *clean* signals, which tend to have a higher overall probability since they all lead to the same choice of which element to try to make, and low average output since there are no informative signals guiding choices. Consistent with the skewed element output distribution, we expect and see a tail to the output distribution at the right, with non-negligible probability, 2.0 %, of output above 6.2. Overall, the output distribution fits our intuition about innovation and creative products, but here expressed at the field level: a relatively small percentage of paths of development have high output, while many paths for the field have relatively low output through five periods. Appendix Figure A2 shows additional results.

Figure 6c, d present statistics on the distributions for height as well as element of greatest length, a natural measure of complexity of created elements. Height is measured as the longest chain from an initial element. As shown in Fig. 6c, nearly all structures cluster in the 4 to 6 height region, with a handful having height 7 and 2 having heights below 4. The complexity distribution in Fig. 6d is considerably more spread out. This is due to the fact that elements are created by overlapping parents of various lengths, and thus while the longest element created is in general the element at the bottom of a chain, these elements vary in length because their parents vary in length. The modal longest length is 5, associated in particular with *clean* signal paths. There are also significant probabilities of structures associated with lengths 6, 9, 12, and even 15, with over 5 % of structures having longest element of length 12. Considering that the field begins with initial elements of heights 2 and 3, this shows a very substantial growth in complexity over 5 periods. Figure 6e presents the distribution of number of initial elements used to create new elements in the field through the first five periods. Far and away, the most common case is for three of the four initial elements to have been used, with the structures associated with this outcome having probability of 0.92.[22]

---

[22]There are 15 structures out of the 131 that use fewer elements, including one case in which no new elements are created. There are 16 structures that use all 4 elements, however, having quite low cumulative probability: these are associated with paths having unusual positive signals that lead to unusual choices for new elements to attempt to make. This is another indication of how important the intuitive signals are for generating diversity in the field, in this case leading to every initial element being used at least once in the creation of a new element, which does not occur for the *clean* paths.

Clean Runs In order to show the importance of intuitive signals in the development of the field, I have developed a set of comparison simulation results, starting from the same initial conditions as above, for the case in which individuals gain no intuitive signals—*clean* runs. In this case, there are typically many fewer paths, indeed for many nodes just a single optimal path; for some nodes, more than one seed has the same expected value and in those cases I assume each seed is chosen with equal probability thus generating more than one path. For this particular simulation, there are 21 clean paths, but they all lead to the same final structure through period five, the first field structure depicted in Fig. 5. The output from this path is 1.2, well below the 2.1 mean value for the signal case, showing the value of information associated with the signals. The number of elements created is 4, slightly below the mean of 4.3 for the signal case. It is intuitive that the value of the signals is expressed more fully in output, since that is what individuals aim to maximize. In particular, in some cases, a signal of a potentially high value may lead an individual to attempt to make an element for which the probability of viability of the main element or associated subbundles is actually somewhat lower, but more than offset by the higher expected output if the element (and subbundle(s)) turn out to be viable. Overall, the clean run shows how much richer the potential development of the field is with the intuitive signals, both in terms of the range of possible paths as well as the number of elements, output, and complexity of created elements and structure. Indeed, an important finding is that it is the signals that generate the very great range of possible paths of development and structures for the field.

Output Dynamics One important calculation is the evaluation of expected output on path segments for which a new element was successfully created in the preceding period, compared with segments for which no new element was created. This is a time series property exploring the correlation in output from one period to the next. There are in fact two offsetting effects that enter into this relationship. When a new element is created, it opens up new frontiers for new seed choices and new elements that it is now possible to try to create. These factors tend to increase expected output, but by how much varies depending on the state of the field. Offsetting this, when a new element is created, it is no longer available to attempt; this effect tends to reduce expected output going forward (recall that which elements are viable is fixed and prespecified in the *masterlist*).[23]

The first row of Table 1 provides information about this tradeoff averaged over all simulation path segments. Conditional on a new element being created in the preceding period expected output is 0.363, whereas conditional on no new element having been created expected output is .399. Thus, the second effect on average is actually

---

[23]To clarify, suppose for the given node the pool of potential elements to attempt to make includes m elements that are viable and n that are not. Suppose further that on one path from the node an element is attempted that is not viable and on a second path an element is attempted that is viable. If in the next period an element is attempted from the remaining elements in the pool (which will be the case for path 1, does not have to be the case for path 2 but may be), then the probability the element will be viable is, other things equal, $n/(n+m-1)$ on the first path and $(n-1)/(n+m-1)$ on the second path, thus greater on path 1.

Table 1 Conditional output based on previous period

	No elements created last period		Element created last period	
	Expected Output	Number of Paths	Expected Output	Number of Paths
All paths	.399	4133	.363	1882
Clean paths	.608	789	.330	426
Other paths	.357	3344	.372	1456

more important, which may seem surprising. However, a further cut of the results reveals an additional factor. It stands to reason that the second, negative effect will be more important when that element was attempted and created based on the individual making the same choices as for the *clean* run, which an individual follows whenever the signals he receives are not useful given the current state of the field. In contrast, when we restrict to paths for which the individual in the preceding segment received at least one active signal, the second effect will typically be less important, since he may well attempt to make an element that would not commonly be attempted. Based on this insight, the second part of Table 1 shows results for expected output conditional on whether or not a new element was created in the preceding period, split by whether the individual in the preceding period followed a *clean* path.[24] We see that conditional on a clean path, expected output is 0.608 when no new element was created in the preceding period, and 0.33 when a new element was created. This large difference highlights the importance of the second effect for clean path segments. In contrast, the bottom line shows that conditional on a path that is not clean expected output is 0.357 when no new element was created in the preceding period and 0.372 when a new element was created. Thus in this case, expected output is greater when a new element was created in the preceding period, though not by a large amount. These results provide empirical predictions about the development of fields: In general, it is not correct that having a success immediately prior to the next attempt in the field is associated with higher output, but that is the case if the preceding attempt was based on active signals and thus entailed attempting to make a more unusual new element.

Path Dependence One important aspect of the diversity of paths of development of the field is the possibility of a high degree of path dependence. Once two paths diverge, there is a high likelihood they will remain distinct and indeed grow apart. This is most evident when along one path, due to an unusual signal, an element is attempted and created and added to the field that is not typically attempted. For once this element is added, it provides a springboard for different seed choices and different intuitive signals being generated, which in turn leads to different elements that

---

[24]This path is the one adopted by individuals who receive signals that are both clean. It may also be followed for other signal draws.

can be and at least in some cases are attempted and created, tending to drive the field even further away from other paths. Table 2 provides evidence on path dependence for the simulation run being discussed. It shows the overlap and diversity between pairs of sets of elements created through period 5 for the 7 period 1 nodes. Node 1 is associated with the *clean* signal path for period 1, as well as a few additional signal paths. It has 29 new elements created over all paths emanating from it. Recall that 3 of the 7 nodes—not node 1—had a new element successfully created, and one, node 2, had more paths and more distinct elements in the subsequent period 3. The numbers in the table corroborate that node 2 is associated with a large, distinct set of created elements all the way out through period 5; there are in total 50 new elements created along all paths emanating from this node. As an example of overlap and diversity, nodes 1 and 2 share 24 elements in common, while node 1 has five new elements associated with its paths that are not created by any path emanating from node 2, and node 2 has 26 elements associated with its paths that are not created along any path emanating from node 1. Interestingly, no pair has identical created sets. Furthermore, in all but two cases, each node set contains elements not contained in the other node set, so that the development of the field in each case produces elements not created along the other path. Thus, there is a high degree of path dependence beginning from the first period.

Learning Sets The simulation results also provide information about learning sets, another area of empirical prediction of the model having policy implications for education. One important issue is how a central administrator might structure learning, as opposed to the learning sets individuals will freely choose. For considering policy around this issue, it makes sense to assume that central administrators have all public knowledge about the state of the field at a given node, including the full history

Table 2 Path dependence: pairwise comparison of Gen 1 node sets

Path Dependence: Pairwise Comparison of Gen 1 Node Sets

Period 1 Node

Period 1 Node	2	3	4	5	6	7
1 29 Els	55, 24, 5, 26	35, 19, 10, 6	34, 29, 5, 0	36, 10, 19, 7	33, 29, 0, 4	41, 18, 11, 12
2 50 Els		57, 18, 32, 7	56, 28, 22, 6	55, 12, 38, 5	56, 27, 23, 6	61, 19, 31, 11
3 25 Els			36, 23, 2, 11	30, 12, 13, 5	35, 23, 2, 10	33, 22, 3, 8
4 34 Els				39, 12, 22, 5	35, 32, 2, 1	42, 22, 12, 8
5 17 Els					39, 11, 6, 22	32, 15, 2, 15
6 33 Els						41, 22, 11, 8
7 30 Els						

Notes: Number below each node is number of elements created along all paths from that node.

In each cell 4 numbers are shown: (i) Number of elements in the union; (ii) number in the intersection; (iii) number in the row node set and not the column node set; and (iv) number in the column node set and not the row node set.

of the field, but do not have any information about the intuitive signals individuals working in the field receive.

Based on this assumption, there is a crucial distinction between seed and full learning sets. For a given node, the optimal seed learning set is based only on public information, thus a central administrator should be able to identify it. However, optimal full learning sets depend on the intuitive signals individuals receive. An administrator, acting only based on public information, will specify a single full learning set at a given node, typically the one compatible with the optimal *clean* choice. When there is more than one optimal full learning set, the administrator will therefore miss opportunities, for some signal pairs, to guide learning optimally. As it turn out, there is significant variation in full learning sets (note that at any node for which the field includes only the initial elements there is no choice regarding the full learning set since all four elements will be included). For period 1, there is only one possible full learning set. For period 2, there are three nodes (out of seven) for which one or more elements were created in period 1 and therefore there is a choice regarding the full learning set. For two of these three nodes, there is more than one optimal full learning set, though in each case, only one out of several paths (six for one node, seven for the other) is different from the other paths and has relatively small probability. In period 3, there are 23 nodes at which the field includes additional elements. For 13 of these, there is more than one optimal full learning set depending on the signals. For 4 of these nodes, there are 3 distinct full learning sets, for the other 9 nodes there are 2. For period 4, out of 161 nodes for which the field includes at least one additional element beyond the initial elements, there are 79 nodes for which there is a single full learning set common to all paths, representing 49 % of the nodes, 32 for which there are 2 distinct full learning sets, 48 for which there are 3, and 2 for which there are 4. Lastly, for period 5, out of 1047 nodes for which the field includes at least one additional element, there are 471 nodes for which there is a single full learning set, representing 45 % of the total number of nodes, 298 nodes for which there are 2 distinct full learning sets, 247 for which there are 3, 29 for which there are 4, and 2 for which there are 5. Thus as the field grows in complexity, there is in general more diversity in full learning sets, representing more than half of all paths in the last two periods.

Full Set of Simulations Table 3 summarizes results for the full set of nine simulations for the base case. Three simulations were performed for each of three initial conditions: $N = 3$ above, $N = 4$, and $N = 8$; each simulation is based on a *masterlist* generated for that simulation. Because the $N = 3$ initial condition includes one additional element (beyond the simple ring), results for this case are slightly less comparable with the other two N values, whereas the $N = 4$ and $N = 8$ cases are more directly comparable. The table shows that the number of paths is substantially higher for the $N = 8$ simulations. Extrapolating specifically from the $N = 4$ results, we see that the number of possible paths of development of the field increases, which is not surprising since there are more basic elements hence more possible distinct combinations. However, the number of distinct field structures through five periods is not significantly higher for the $N = 8$ case than for $N = 4$ or even $N = 3$—there are more duplicate paths.

Table 3 Full set of base scenario simulations: summary statistics

	Scenario								
	$N = 3$			$N = 4$			$N = 8$		
	Sim 1	Sim 2	Sim 3	Sim 1	Sim 2	Sim 3	Sim 1	Sim 2	Sim 3
Number of paths	6015	5958	5511	3694	1246	2725	20,400	14,660	17,478
Number of Distinct									
Structures	131	20	111	136	5	48	92	9	143
Expected height	4.75	5.43	4.68	4.84	3.00	3.98	4.29	3.05	3.90
Expected max length	6.55	7.50	8.96	6.97	4.00	5.06	5.59	4.07	5.00
Expected Number									
of elements	4.24	4.25	5.63	5.68	1.72	2.50	5.27	2.05	3.59
CLEAN expected									
number of elements	4.00	1.25	3.00	3.58	1.00	1.60	4.81	1.55	2.34
Expected output	2.10	1.72	1.77	1.66	.505	.732	1.69	.627	4.46
CLEAN expected output	1.20	.345	.875	1.36	.323	.490	1.65	.475	.786
Conditional Output:									
All paths: 0	.399	.371	.286	.264	.089	.143	.413	.139	.689
1	.363	.495	.328	.210	0.0	.196	.400	.044	.926
Conditional output:									
CLEAN paths: 0	.608	.122	.240	.316	.102	.201	.355	1.69	.639
1	.330	.213	.300	.141	0.0	.140	.411	.012	.191
Conditional output:									
Not clean paths: 0	.357	.402	.325	.250	.082	.117	.439	.120	.742
1	.372	.683	.365	.238	0.0	.211	.389	.081	1.23

It is noteworthy that the number of distinct structures varies very widely, from a low of only 5 to a high of 143. Indeed, the distribution has two main regions, one containing runs with quite low numbers of distinct structures, 20 or below, and a second region containing runs for which there is a far larger number of distinct structures, above 90; only 1 out of 9 simulations does not fall in these two regions. Thus the degree of variability in set of possible paths of development of a field is itself highly variable.

A consistent regularity across all simulations is that expected output is higher for the signal model than for the clean model, reflecting the value of information in the signals. The differential between the two varies substantially, reflecting the high variability in output, especially due to the right tail of the output distribution. While the expected number of elements would not necessarily be greater for the signal model, since as noted above it is expected output that is being maximized not expected number of new elements formed, nonetheless it is also consistently higher. Conditional expected output is higher following a successful new element being made in four out of the nine simulations, and higher following no new element being made for

Table 4 High viability probability: summary statistics

	Sim 1	Sim 2	Sim 3
Number of paths	10,548	5531	2679
Number of distinct structures	1408	90	409
Expected height	6.21	4.22	4.54
Expected max length	10.7	5.27	6.64
Expected number of elements	13.8	1.70	5.71
CLEAN expected number of elements	9.86	1.25	6.33
Expected output	4.15	.574	1.76
CLEAN expected output	3.52	.417	1.83
Conditional output: All paths: 0	1.00	.131	.516
1	1.25	.311	.515
Conditional output: CLEAN paths: 0	.926	.171	.583
1	1.20	.237	.540
Conditional output: Not clean paths: 0	1.01	.123	.488
1	1.27	.328	.510

the remaining five simulations. Differentiating between cases when a clean path has been followed, versus a signal path, conditional expected output is higher following a new element being made in three out of nine simulations for clean paths, and in five out of nine simulations for signal paths. Furthermore, the difference in expected output between when a new element is formed and when it is not is greater for the signal paths in seven out of the nine simulations. Thus, the intuition discussed above is largely born out, that making an existing element is a more important negative factor, relative to the benefit of opening up new possibilities with the new element, for clean paths than not clean paths.

Figures A3 and A4 in the Appendix provide additional results.

Alternative Scenario: High Viability Probability I ran an alternative scenario for $N = 3$ altering the probability a new element is viable from 0.3 to 0.5. We expect more new elements to be produced, and the simulations of this scenario show how this translates into differences in the way the field may develop. Table 4 presents summary statistics for three simulation runs for this scenario.

The most striking difference between these results and those for the baseline scenario is that there are many more distinct structures. This is related to the fact that

there is a greater diversity of created elements. There are two reasons for these results. One is that more elements that are attempted are made, which leads to a larger set of made new elements each period.[25] The other is the fact that, with a higher baseline probability of viability, individuals are free to respond more to the probability a given element will have a value drawn from the high distribution in making their choice of which element to attempt to make. Being more sensitive to these signals leads to a greater variety of choices of which elements to attempt along different paths. A further result is that there is now a somewhat larger, though still small, chance of a very high output draw. An illustration of this is the output distribution for the first simulation shown in Appendix Figure A5: there is a very small probability, associated with just two paths, of a very high output above 100.0. Finally, a third interesting result is that expected output and number of elements is relatively higher for the clean paths, compared with the baseline case. This is because the value of information associated with the viability signals is greater when the baseline probability is lower.[26] Overall fields with higher baseline probability of viability can be expected to exhibit greater variety of created elements and a small but non-negligible probability of a very high output path.

Conclusion

In this paper, I have presented a formal model of the creative development of a field. The heart of the model is an individual rational learning process leading to the generation of new elements in the field. The output of the model describes how the field grows, with a structure resembling a lattice. The model is used to explore a range of issues, including the diversity of possible paths of development of the field, how individuals build on the work of their predecessors, and the role of intuitive signals in the development of the field. Main results include the very substantial diversity of generated structures for the field's development, the importance of intuitive signals both in guiding creative development and generating the diversity of outcomes, and a high degree of path dependence. The model generates implications about output, including the variance of output and serial correlation in output.

The model can be extended in a number of important ways. One extension I have explored incorporates intellectual property (IP) protection into the model. In this extension, individuals are paid a royalty when an element they create is used in subsequent periods. I have specifically explored this model for the case in which royalty payments extend for 1 period and the royalty payment is a percentage of

---

[25]Note that this does not extend to the number of paths being larger. Paths are generated even when an element is not viable, since the history update is distinct.

[26]Reflecting the lower value of information of signals, in the third simulation, the expected number of elements and output are actually higher for the clean run than the full model, a result which does not obtain in any of the nine baseline simulations. This is possible when there are some elements attempted along a clean run that are not attempted for the signals; whether or not these elements are viable and their output is drawn at random and thus a high proportion may by chance turn out to be viable and have high value.

the total value. We expect that a royalty policy of this kind will encourage individuals to focus on producing elements that will be useful as building blocks by later entrants into the field. Preliminary work with the royalty set at 30 % of value shows this to be the case but the overall effect on the development of the field is modest. In particular and as an example, replication of the base scenario produces the following results. There is no difference in paths of development in period 1. In period 2, five new paths are generated within the three nodes that have a new element created from period 1. However, the overall probability of these new paths is modest, just a few percent with output increased by approximately 1 % overall. A second important extension of the model is for more than one individual to work in the field each period. This is obviously important for exploring the effects of competition on learning and production choices. Lastly, team production can be explored, such that individuals learn separately and then are given the opportunity to combine their knowledge.

It is also possible to extend this model in ways that may link the approach here with other approaches or models of the creative process. Thus, the role of extrinsic and intrinsic motivation on creativity can be studied. Extrinsic motivation might include a royalty payment, discussed above, payments for learning and sharing information, and payments to reveal intuitive signals. Intrinsic motivation would presumably involving modeling the psychic cost of learning and exploring how the model predictions change as this cost changes. In addition, the model here is one of simple expected value maximization and an extension could explore the role of risk influencing creative development. Furthermore, it would be interesting to compare the predictions of this model with those of the very simple Darwinian model pioneered by Campbell.

As a second area of development, the model might be linked with higher level institutional, economic, and sociological approaches. In the model here, individuals learn on their own. With a cohort of individuals, we may explore an educational system in which individuals are taught a core curriculum as a group and then pursue independent learning (see Feinstein 2015). Furthermore, individuals may work in institutional settings and this raises the question of the allocation of resources across institutions and individuals to support learning paths and creative projects. Finally, from a behavioral perspective, it would be interesting to explore the role of popularity and fads in individuals' learning and project choices. For example, perhaps "hot topics" are more likely to spark intuitive signals, which might add to the dynamics of how individuals build on previous work, perhaps alternating dynamics so that individuals are more likely to build on recent work.

In general, the approach taken in this paper opens the way to modeling knowledge creation in a more structured way than has been done previously in formal decision-based models of creativity and innovation. This forges an important bridge between knowledge representation and economics, centering on creativity and innovation and the production of new knowledge. This is a link truly in its infancy with a great deal more to be done if we are to represent the incredibly rich knowledge sets individuals form, the distinctiveness of knowledge sets across individuals, and the importance of these sets for creativity and innovation in all its forms, including strategy, technology, the arts and sciences, and policy. More broadly, the structure of knowledge

is integral to the human condition, including the modern economy and human culture. Only by representing this structure can we hope to be able to understand how we as individuals build on the work of our predecessors, in specific conceptual ways and following defined rational decision processes, and how humans in general respond to circumstances, including shocks to their environment, in innovative ways.

References

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–51.

Aghion, P., Harris, C., Howitt, P., & Vickers, J. (2001). Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies*, *68*(3), 467–92.

Amabile, T.M. (1996). *Creativity in context*. Westview Press.

Barron, F. (1969). *Creative person and creative process*. Reinhart & Winston: Holt.

Bramoullé, Y., & Saint-Paul, G. (2010). Research cycles. *Journal of Economic Theory*, *145*, 1890–1920.

Campbell, A., & Feinstein, J.S. (2013). What is taught in core micro courses at top US doctoral programs: overlap and diversity. working paper.

Campbell, D. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400.

Cohen, L.M. (2009). Linear and network trajectories in creative lives: A case study of Walter and Roberto Burle Marx. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(4), 238–48.

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In Sternberg, R.J. (Ed.), *The nature of creativity* (pp. 325–339). Cambridge University Press.

de Bono, E. (1970). *Lateral thinking.* Harper & Row.

Dorf, R.C., & Svoboda, J.A. (2006). *Introduction to electric circuits*, 7th Edn. Wiley.

Evenson, R.E., & Kislev, Y. (1976). A stochastic model of applied research. *Journal of Political Economy*, *84*(2), 265–282.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, *22*(2), 133–187.

Feinstein, J.S. (2006). *The nature of creative development*. Stanford: Stanford University Press.

Feinstein, J.S. (2011). Optimal learning patterns for creativity generation in a field. *American Economic Review Papers and Proceedings*, *101*(3), 227–232.

Feinstein, J.S. (2015). Creative development: Patterns of learning. *IMCIC Papers & Proceedings*, 32–37.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*, 909–928.

Florida, R.L. (2002). *The rise of the creative class*. Basic Books.

Gabora, L. (2005). Creative thought as a non Darwinian evolutionary process. *The Journal of Creative Behavior*, *39*(4), 262–282.

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of European patents. *European Management Review*, *5*, 69–84.

Ganter, B., Stummer, G., & Wille, R. (2005). *Formal concept analysis: Foundations and applications*. Berlin: Springer.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, *108*(5), 874–904.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gruber, H. (1974). *Darwin on man: A psychological study of scientific creativity*. New York: E.P. Dutton.

Harhoff, D., Scherer, F.M., & Vopel, K. (2005). Exploring the tail of patented invention value distributions. In Harhoff, D., Scherer, F.M., & Vopel, K. (Eds.), *Patents: Economics, policy, and measurement* (pp. 251–81). Cheltenham: Elgar.

Hayek, F.A. (1960). *The constitution of liberty*. Chicago: University of Chicago Press.

Helbig, H. (2006). *Knowledge representation and the semantics of natural language*. Berlin: Springer.

Jaffe, A.B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. Cambridge: MIT Press.

Jackson, M.O. (2008). *Social and economic networks*. Princeton: Princeton University Press.

Jones, B.F. (2009). The burden of knowledge and the "Death of the Renaissance man": Is innovation getting harder? *Review of Economic Studies*, *76*(1), 283–317.

Kaplan, S., & Vakili, K. (2013). Novelty vs. usefulness in innovative breakthroughs: A test using topic modeling of nanotechnology patents. working paper.

Kenett, Y.N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00407.

Keynes, J.M. (1977, 1981). The collected writings of John Maynard Keynes. Volume XVII; Activities, 1920–1922; Treaty revision and reconstruction. In Johnson, E., & Moggridge, D. (Eds.), *Volume XIX; Activities, 1922–1929; The return to gold and industrial activity*. London: Macmillan and Cambridge University Press.

Koestler, A. (1964). *The act of creation*. New York: Macmillan.

Kortum, S.S. (1997). Research, patenting, and technological change. *Econometrica*, *65*(6), 1389–1419.

Lancaster, K.J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157.

Li, M., & Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York: Springer.

Lindell, A.K. (2011). Lateral thinkers are not so laterally minded: Hemispheric asymmetry, interaction and creativity. *Laterality*, *16*(4), 479–498.

Marshall, A. (1919). *Industry and trade*. London: Macmillan.

Mas-Colell, A., Whinston, M., & Green, J. (1995). *Microeconomic theory*. New York: Oxford University Press.

Matisse, H. (1990). In Flam, J.D. (Ed.) *Matisse on art*. New York: Phaidon Press.

Mead, C., & Conway, L. (1980). *Introduction to VLSI systems*. Addison-Wesley.

Mednick, S.A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.

Mill, J.S. (1978 originally published 1859). *On liberty*. Indianapolis: Hackett Publishing.

Mokyr, J. (2002). *The gifts of Athena: historical origins of the knowledge economy*. Princeton: Princeton University Press.

Poincaré, H. (1908, 1952). Mathematical creation. In Ghiselin, B. (Ed.), *The creative process: a symposium* (pp. 33–42). Berkeley: University of California Press.

Scherer, F.A., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, *29*, 559–566.

Simonton, D.K. (2003). Scientific creativity as constrained stochastic behavior: the integration of product, process, and person perspectives. *Psychological Bulletin*, *129*, 475–494.

Silverberg, G., & Verspagen, B. (2007). The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, *139*, 318–339.

Skidelsky, R. (1983). *John Maynard Keynes. Volume two; The economist as saviour*. London: MacMillan.

Solow, R.M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics*, *39*(3), 312–320.

Sowa, J.F. (1984). *Conceptual structures: information processing in mind and machine*. Reading: Addison-Wesley.

Sowa, J.F. (2000). *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove: Brooks/Cole.

Spurling, H. (1998). *The unknown matisse*. Berkeley: University of California Press.

Sternberg, R.J. (1999). *Handbook of creativity*. Cambridge: Cambridge University Press.

Torrance, E.P., & Ball, O.E. (1984). *Torrance tests of creative thinking (Revised)*. Scholastic Testing Services.

Weitzman, M.L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360.

Wille, R. (1992). Concept lattices and conceptual knowledge systems. *Computers and Mathematics with Applications*, *23*, 493–515.