

The Distinctiveness of Individuals: Using Knowledge Structures To Represent Research Focuses; The Field of Stem Cell Research

Jonathan S. Feinstein
Yale School of Management

1. Introduction

A fundamental challenge for economics and the social sciences is the development of models of creativity and innovative activity, spurring innovation and cultural development, across all fields. Originality is rooted in distinctiveness: an individual comes to make original, even unique contributions because in some way he or she is different from others and has different experiences, thus is able to create or invent something or some concept that others cannot create or invent.

To trace the creative, inventive process in these terms means, as I take it, representing the distinctiveness of individuals' creative interests and research focuses, as well as the distinctive paths of development they follow – experiences they have, phenomena they encounter and observe, events they witness, and elements they encounter, including through reading books and articles, conversations, and attending social and cultural events – through which they come to make their original contributions. In turn, developing such representations requires moving beyond models in which all individuals are essentially interchangeable, with very similar profiles differing only in a few parameters, an approach common to both neoclassical and behavioral economic models – for in such models it is not possible to represent, at least not in a deeper more realistic way, the ways in which individuals differ from one another that lead them to make distinct, unique creative contributions.

In this brief article I outline a methodology for addressing this challenge. I describe development of a methodology describing the research focuses of interest of a set of individuals working in a field of creative endeavor, which could be a scientific or engineering field, a business sector, for example financial engineering or product development in a particular industry, or a field of art or design. The approach I describe is based on building a knowledge structure representing the field, specifically the set of conceptual and practical elements in the field, and representing the focuses of interest and research activities of individuals and work groups in terms of the knowledge structure. This representation can then be used to study the differences in focuses among individu-

als working in the field, and the specific elements and combinations of elements that make each individual's or work group's focus distinctive.

In *The Nature of Creative Development* [1] I describe creative development: I show that individuals creative interests are at the heart of their creative development and work, and describe a set of specific pathways through which individuals, through pursuing and developing their interests creatively, generate ideas, insights, and creative projects, leading to creative contributions and innovations. The methodology I describe here weds the model of creative development in my book with a framework for knowledge representation as developed and described in computer science, artificial intelligence, and the representation of conceptual structures, as for example presented by Sowa [2,3].

The methodology I present shows a way leading on to the development of models describing individuals' rich conceptual worlds and the ways their rich conceptual worlds are the basis for their creative and professional activities and contributions. Such modeling is fundamentally at odds with the extreme simplification of the human conceptual world that is common to nearly all modeling at present in the social and behavioral sciences. Standard modeling approaches do not represent the richness of individuals' conceptual worlds sufficiently to manifest the ways in which each individual's conceptual world is distinctive – which is the essential basis for creativity and innovation.

I mention as an application ongoing work developing a representation of research focuses of researchers working in the field of human embryonic stem cell (HESC) biology. For this work we are developing a knowledge structure of the field of HESC biology; we follow the general approach developed by the Gene Ontology Consortium for mapping the knowledge structure in cellular and molecular biology [4,5].

2. Field Knowledge Structures and the Representation of Research Focuses

A field is a culturally constituted domain in which individuals work – so that main activities they engage

in, specifically knowledge acquisition and work practices, center and are based in the field. (A field may also be defined in terms of sociological and economic terms, for example as a domain for resource allocation and competition for prestige, but I abstract from those issues here.) Examples of fields are scientific fields, fields of artistic endeavor, fields of engineering, product development and design fields, and practices in which there is ongoing development of knowledge and techniques, including medicine, law, finance, and management.

I focus on representation of a field as a knowledge structure (KS). Ideally the KS representing a field should contain the full set of concepts, topics and subtopics, approaches, facts, results, hypotheses, practices, and individual works and elements that constitute the field. This structure can be organized as a graph such that two elements are connected if they are related in the context of the field. A basic form of relationship is parent-child. An example in stem cell biology is: the category “stem cell” is parent and “embryonic stem cell” is its child. A given element may have multiple parents and children, and two or more elements may share children. “Stem cell” for example has at least two children, “embryonic stem cell” and “adult stem cell.”

More complex concepts are also defined in the KS. For HESC biology, a fundamental process included in the KS is “cell fate determination.” Another important concept is the day post coitus (dpc) when a cell exists or a process or event occurs. The KS also includes other kinds of conceptual links among elements, in addition to parent-child, defining processes, activities, and products connecting elements. For HESC an important link is between “cell fate determination” and “embryonic stem cell,” with distinct links at distinct dpc dates; each such link is in turn connected with a specific cell-type that emerges in the embryo at that time, such as “neural progenitor cell.” A second kind of link is a link of two elements that often occur together in work activity in the field or in biologic processes or end products. An example is the link of a specific *in vitro* method with a specific nutrient typically used in the *in vitro* preparation; in turn this pairing links to the specific kinds of HESCs that are supported.

A knowledge structure defined solely by parent-child relationships may be represented as a directed acyclic graph or *dag* (it is not in general a lattice, as the meet or join of a pair of elements may contain multiple elements). As I have described, a

KS includes other relationships in addition to parent-child. Thus it is not typically formally a *dag* – it may be thought of as a *dag* with supplemental connections.

The KS for a field is likely to be a quite large structure. The gene ontology KS currently contains approximately 20,000 elements. For stem cell biology we expect our KS to be significantly smaller, but still to number in the hundreds of elements, with several distinct hierarchies, and with hierarchies having multiple levels.

Each individual working in a field forms his or her own mental representation of the field, essentially his personal KS that depicts the field as he perceives it. These will inevitably differ across individuals – as individuals elaborate certain subdomains more than others, and conceptualize different relationships among elements. Ideally it thus would be appropriate to elicit from each individual working in a field his own KS representation of the field. This would be difficult to do, indeed not possible in any complete way, as eliciting an entire KS representation from an individual is a large undertaking (however, it might be possible to elicit top and mid-level portions of such structures). In our HESC project we strive to recognize some differences among people, for in fact definitions of certain critical stem cell biology terms are at present contested. However, we focus on building a KS that is in a sense taken to be common knowledge and to a first approximation agreed upon by individuals working in the field; we then build alternative structures that deviate from the initial structure in specific details. Building a KS that is at least to a degree shared by individuals knowledgeable about the field is valuable in that the KS is objective and can be defined in a clear manner. However, a focus too narrowly on a single KS can mean that an important aspect of distinctiveness of individuals is lost. Thus we strive for some balance. Even in the case of a single KS distinctiveness can still be represented, in terms of the different specific locations on the KS of the distinct focuses of interest of different individuals, as defined by the sets of elements that form the bases for their descriptions of their focuses of interest, and that is the approach we pursue.

The notion of a focus of research is based on the concepts of creative interests and conceptions of creative interests developed in *The Nature of Creative Development* (see especially Chapter 2). A creative interest an individual has is a topic, typically of his own construction, that he finds fascinating, that he is

curious about and interested in exploring and learning about, that he intends or at least hopes to pursue himself – to develop creatively, to employ as the basis or center for creative work in his field. A creative interest may be a substantive topic or subject, or alternatively may center on an approach, a set of principles, or an objective defined in a broad way, not in precise detail. A conception of a creative interest is a conception an individual forms of his interest, defining it and imagining it, which is the basis for descriptions he provides of his interest to others. Creative interests possess two critical characteristics: they are of intermediate breadth, and are distinctive to the individual, thus a root of his originality.

A research focus is a researcher’s conception of his main active creative interests; it includes current projects rooted in his interests, and perhaps objectives for what he hopes ultimately may be attained through pursuit of his interests. A main hypothesis I wish to demonstrate is that individuals’ focuses of research are distinctive. By formally representing focuses of research using the KS of a field it is possible to depict this distinctiveness and develop quantitative measures of distinctiveness.

We expect a research focus of interest to be generative of creative projects and creativity through several routes; I describe these pathways in *The Nature of Creative Development*. First, the researcher is guided in his decision-making about what projects to initiate, including collaborative projects, by his conception of his research focus, specifically the fit of potential projects with his focus. Second, a researcher’s focus acts like a filter, guiding his attention and learning. In this way his focus sparks creative responses he has, in which a stimulus he encounters – for example, a recent paper or talk he attends, or phenomenon he observes – sparks an idea, including an idea for a project. Finally, his focus guides his exploration, through which he builds up knowledge in the domain of his interest and connected with it. In turn, this accumulated knowledge may spark ideas and insights, for example the recognition of a general pattern or a creative connection. I present many examples of both creative responses and creativity rooted in knowledge built up through exploring interests in *The Nature of Creative Development*.

In the present study, we will explore how the distinctiveness of researchers’ focuses of interest is the basis for and reflected in their subsequent ideas and projects, generating these through the pathways described above. Specifically, using the KS we can show

how the distinctiveness of individuals’ focuses of research is conceptually linked to their subsequent ideas and projects by identifying specific distinctive conceptual elements or sets of elements in their focuses that are either incorporated in their ideas and projects or linked, via the KS or more generally, to distinctive elements in their ideas and projects. This hypothesis can be explored only in the fullness of time, over a several year period, as individuals pursue their interests and generate ideas and projects leading to defined contributions. Supplemental issues to explore may include the relationship of focuses of research with years since earning a degree, gender, and dissertation advisor.

3. Application: Human Embryonic Stem Cell Research

I am engaged in a project (with research assistant Lauren Ro) collecting information about researchers in the field of human embryonic stem cells (HESCs) [6,7]. Our goal is to “map” the HESC field – develop a KS defining the field, use the KS to describe the focus of research of each active researcher or research unit in the field, show that researchers have distinct focuses, and quantify the degree of distinctiveness. Note that the HESC field includes not only individuals engaged in biological research but also individuals interested in ethical, social, and economic aspects of HESC research.

In the first stage of our project we are identifying, as far as is possible, the universe of researchers who are or recently have been and may in the near future again be active in the HESC field. To identify researchers we are pursuing several approaches simultaneously. Our main approach is to search the PubMed database for recent articles in the field. We use as our main keyword phrase “human embryonic stem cell”. This phrase returns approximately 100 articles per month. We focus on the period November, 2005 through August, 2006. We scan all articles that are returned and retain those with a clear focus in this field and eliminate those which are not in the field (thus for example articles focusing on murine stem cells or adult stem cells, which may be identified by the keyword search because they mention the words “human” and “embryonic,” are eliminated). We include all individuals who are listed as an author on one or more of these papers in our master file of potential researchers. In addition to searching PubMed we also draw on *Who’s Who in Stem Cell Research* [8] which lists active academic researchers and companies in the field. For nearly all academic researchers we find that anyone included in the *Who’s Who di-*

rectory appears also in the PubMed search; but the reverse is not true, and for companies it is also not the case. Finally, when we contact heads of research units and administrators (see below) we ask for names of individuals in their laboratories or units who are actively engaged in research in their laboratory in the HESC field.

Next we contact or otherwise seek to gain information about all individuals listed in our file of potential researchers, in order to identify researchers currently active in the field or likely to be active in the field in the near future. Individuals who are not active in the field, for example technicians and specialists in other fields and former graduate students who are no longer engaged in research, are eliminated. We rely on administrators and laboratory heads for information about individuals for whom we are not independently able to obtain information about their current position and work. Our file currently contains more than 400 researchers active in the field.

In the second stage of our project we are building a knowledge structure (KS) for the HESC field. We strive to build a KS that is specifically useful for representing researchers' research focuses. This means we include both more traditional conceptual elements that define the field substantively as well as other elements and substructures that are important for individuals in defining their interests and focuses of research, such as laboratory techniques, means of obtaining stem cells, and ethical principles.

In the third stage of our project we are collecting information about researchers' focuses of interest. For this purpose we utilize two main approaches. For individuals who have Webpages, we scan their Webpages for a description of their current focus of research. For individuals who do not have Webpages, or for whom there is no clear statement on their Webpages of their focus of research, we contact the individual, by email or telephone, and seek to elicit from them a statement describing their current focus of research and any focuses they plan or hope to pursue in the future.

Finally, in the fourth stage of our project we are defining individuals' research focuses in terms of concepts in our KS and analyzing these focuses. In particular, we are able to measure the distance, in terms of separation on the KS of sets of elements (there are different possible metrics), between any two focuses of research, and thus develop metrics quantifying the degree of distinctiveness of focuses of research. We also can analyze and depict the universe

of focuses, to show how these cluster in certain regions of the KS, and also to show that each researcher has a focus that is distinctive.

4. Conclusion

In this article I have outlined a new approach for depicting individuals working in a field, developing a conceptual framework based in a KS representation of the field, that enables the distinctiveness of each individual in his creative interests, focuses of activity, projects, and contributions to be described and measured. I have also outlined an application of this approach, an ongoing research project building a knowledge structure and gathering information about focuses of research of researchers in the field of human embryonic stem cells.

It is my hope that development of the methodology outlined here will contribute towards the development of a new approach in the social sciences: to depict individuals' rich conceptual worlds and manifest the ways individuals are distinctive, in their interests, paths of development, and contributions to society.

5. References

- [1] Feinstein, Jonathan S. *The Nature of Creative Development*. Stanford, CA: Stanford University Press, 2006.
- [2] Sowa, John F. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley, 1984.
- [3] Sowa, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole, 2000.
- [4] The Gene Ontology Consortium. "Gene Ontology: tool for the unification of biology." *Nature Genetics*, Vol. 25, May 2000, pp. 25-29.
- [5] Gene Ontology Consortium. "The Gene Ontology (GO) database and informatics resource." *Nucleic Acids Research*, Vol. 32, 2004, pp. D258-D261.
- [6] Lanza, Robert, John Gearhart, Brigid Hogan, Douglas Melton, Roger Pedersen, James Thomson, E. Donnall Thomas, and Michael West, Editors. *Essentials of Stem Cell Biology*. Burlington, MA: Elsevier Academic Press, 2006.
- [7] Sell, Stewart, Editor. *Stem Cells Handbook*. Totowa, NJ: Humana Press, 2004.
- [8] *Who's Who in Stem Cell Research*. Leesburg, VA: DataTrends Publications, 2006.