1-2011

Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas

Laura J. Kornish

Karl Ulrich
University of Pennsylvania

Recommended Citation

Kornish, L. J., & Ulrich, K. (2011). Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas. *Management Science, 57* (1), 107-128. http://dx.doi.org/10.1287/mnsc.1100.1247

Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas

Abstract

A common approach to innovation, *parallel search*, is to identify a large number of opportunities and then to select a subset for further development, with just a few coming to fruition. One potential weakness with parallel search is that it permits repetition. The same, or a similar, idea might be generated multiple times, because parallel exploration processes typically operate without information about the ideas that have already been identified. In this paper we analyze repetition in five data sets comprising 1,368 opportunities and use that analysis to address three questions: (1) When a large number of efforts to generate ideas are conducted in parallel, how likely are the resulting ideas to be redundant? (2) How large are the opportunity spaces? (3) Are the unique ideas more valuable than those similar to many others? The answer to the first question is that although there is clearly some redundancy in the ideas generated by aggregating parallel efforts, this redundancy is quite small in absolute terms in our data, even for a narrowly defined domain. For the second question, we propose a method to extrapolate how many unique ideas would result from an unbounded effort by an unlimited number of comparable idea generators. Applying that method, and for the settings we study, the estimated total number of unique ideas is about one thousand for the most narrowly defined domain and greater than two thousand for the more broadly defined domains. On the third question, we find a positive relationship between the number of similar ideas and idea value: the ideas that are least similar to others are not generally the most valuable ones.

Keywords

search, opportunity, opportunities, idea, ideation, idea generation, innovation, creativity, innovation process, opportunity identification, concept development, product development, product design, entrepreneurship

Disciplines

Business | Finance and Financial Management

OPPORTUNITY SPACES IN INNOVATION: EMPIRICAL ANALYSIS OF LARGE SAMPLES OF IDEAS

Laura J. Kornish
Leeds School of Business
University of Colorado
UCB 419
Boulder, CO 80309 USA
kornish@colorado.edu

Karl T. Ulrich
The Wharton School
University of Pennsylvania
500 Huntsman Hall
Philadelphia, PA 19104 USA
ulrich@wharton.upenn.edu

Original Version: November 2009
This Version: July 2010

Abstract

A common approach to innovation, *parallel search*, is to identify a large number of opportunities and then to select a subset for further development, with just a few coming to fruition. One potential weakness with parallel search is that it permits repetition. The same, or a similar, idea might be generated multiple times, as parallel exploration processes typically operate without information about the ideas that have already been identified. In this paper we analyze repetition in five data sets comprising 1,368 opportunities and use that analysis to address three questions: (1) When a large number of efforts to generate ideas are conducted in parallel, how likely are the resulting ideas to be redundant? (2) How large are the opportunity spaces? (3) Are the unique ideas more valuable than those similar to many others? The answer to the first question is that while there is clearly some redundancy in the ideas generated by aggregating parallel efforts, this redundancy is quite small in absolute terms in our data, even for a narrowly defined domain. For the second question, we propose a method to extrapolate how many unique ideas would result from an unbounded effort by an unlimited number of comparable idea generators. Applying that method, and for the settings we study, the estimated total number of unique ideas is about one-thousand for the most narrowly defined domain and greater than two-thousand for the more broadly defined domains. On the third question, we find a positive relationship between the number of similar ideas and idea value: the ideas that are least similar to others are not generally the most valuable ones.

Keywords: search, opportunity, opportunities, idea, ideation, idea generation, innovation, creativity, innovation process, opportunity identification, concept development, product development, product design, entrepreneurship

1. Introduction

A common approach to innovation is to identify a large number of opportunities and then to select a subset for further development, with just a few coming to fruition. We define *opportunity* as an idea for an innovation that may have value after further investment of resources. For example, in the movie industry an opportunity is a script summary; in the pharmaceutical industry, an opportunity is a newly discovered chemical compound; for an entrepreneur, an opportunity is an "idea…for [a] potentially profitable new business venture…." (Baron and Ensley, 2006).

Hundreds or thousands of opportunities may be considered for every commercial success (Stevens and Burley 1997). This process can be thought of as a tournament of ideas (Terwiesch and Ulrich, 2009), in which many ideas are explored in parallel with only the best prevailing. The parallel-search tournament is one of the standard approaches to exploring a space of opportunities (Sommer and Loch 2004).

One potential weakness with parallel search is that it permits repetition. The same, or a similar, idea might be generated multiple times, as parallel exploration processes typically operate without information about the ideas that have already been identified. (For ease of exposition, we use the terms *idea* and *opportunity* interchangeably.) In practice, repetition might be dismissed as an unavoidable nuisance. In this paper we quantify the extent of repetition in five data sets and show how the repetition provides valuable clues about the size of the opportunity space.

To our knowledge, no prior research has measured or analyzed repetition in opportunity identification. The existing literature either assumes that the identified opportunities are each unique (e.g., Dahan and Mendelson, 2001) or focuses on search strategies over stylized landscapes (e.g., the NK models). In contrast, we explicitly allow for repetition, measure it empirically, and examine its implications. Our goal is to answer both fundamental scientific questions about opportunity identification and to inform managerial practice. This research is motivated by three key questions.

1. How much redundancy results from parallel search? To the extent that there is redundancy, the identification of the same idea multiple times, investments in opportunity

identification are wasted. Answering this question is critical to deciding how much to invest in parallel search.

2. How large are opportunity spaces? Once we know the level of redundancy, we have a clue to the effective size of the opportunity space, the total number of unique ideas. An innovator who has generated 50 unique opportunities would benefit from knowing if there are 100 or 1000 more opportunities to be discovered.[1] In this paper we develop a method for estimating the size of opportunity spaces. This method can be used to find the total number of unique ideas or to find the total number of themes or "neighborhoods" of ideas.

3. Are unique ideas, i.e., those that are similar to no or few other ideas in the data set, more valuable than ideas that are similar to many others? To answer this question, first we establish that sets of generated ideas, do, in fact, show significant clustering, compared to a random benchmark. Then, we test the hypothesis that unique ideas or those found in smaller clusters are more valuable than ideas found in larger clusters.

To address these questions, we analyze a total of 1,368 ideas from five data sets, each created by different groups of individuals who generated ideas in parallel. Our results show that in the data sets we analyze, strict redundancy is not highly prevalent, which suggests that the opportunity spaces are large, on the order of thousands of opportunities. Although strict redundancy is not widespread, we can clearly identify clusters of similar ideas. Our results suggest that cluster size is a positive indicator of the value of ideas. Furthermore, identifying themes for clusters can itself be a useful step in an innovation process, creating a map of the innovation landscape.

The paper is organized as follows. First we discuss prior research in related areas. Then we present a population model for estimating the size of an opportunity space. Next we describe our data and metrics. Then, we describe our analyses in detail and report our results. Finally, we discuss the results and their implications for practice, qualify our findings, and provide concluding remarks.

---

[1] One could argue that the number of ideas is infinite because a detail can always be tweaked to make a new idea or because ideas could be arbitrarily unrelated to the innovation charge. However, the opportunity space can be thought of as finite if ideas that are highly similar are counted together and if ideas that are highly "distant" are assumed to be very unlikely to be generated. We discuss these issues in Section 5 of the paper.

2. Prior Work

This study intersects several rich streams of prior research: (1) creativity and idea generation, (2) models of search strategies, and (3) process models of innovation. Our research also relies on prior work in wildlife ecology and in network analysis. However this reliance is more methodological than conceptual, and so we discuss the literature related to our methods in the analysis section of the paper.

Creativity and Idea Generation

Creativity and idea generation have been examined both in the social psychology literature and in the innovation management literature. The social psychology literature on idea generation originates with the development of *brainstorming* (Osborn 1957). Diehl and Stroebe (1987) and Mullen et al. (1991) provide a detailed overview of this literature. Most studies have experimentally examined groups generating ideas as teams or as individuals. The research has unequivocally found that the number of ideas generated (i.e., productivity) is significantly higher when individuals work by themselves and the average quality of ideas is no different between individual and team processes. (All of these studies normalize for total person-time invested to control for differences in the numbers of participants and the duration of the activity.) These studies have led to prescriptions that idea generation for innovation should include significant efforts by individuals working independently of one other (Ulrich and Eppinger, 2008). This literature provides some of the justification for parallel search in innovation, however that literature does not explicitly address the possibility that parallel search might lead to repetition, a question we address.

The innovation management literature contains large-scale empirical studies of creativity in innovation. Fleming and Mingo (2007) provide an excellent synthesis of the concepts in this literature. These studies often use patent data (e.g., Singh and Fleming (2009), Fleming et al. (2007)), and draw on citations and patent classes to measure relationships among creative ideas (the patents). Fleming et al. (2007) investigate the "size" of an inventor's search space by using a count of subclass combinations. The concept of similarity of ideas is central to our work, and we rely on human raters to make similarity judgments. Part of our contribution is the application of a population model from wildlife ecology to estimate the size of the opportunity space based on the similarity of ideas generated.

Models of Search Strategies

Search is a common paradigm for understanding problem solving generally and innovation more specifically. March and Simon (1958) were among the first to characterize problem solving as search. (See also [Simon 1996].) Subsequently, many scholars have framed innovation as a search problem, including Stuart and Polodny (1996), Martin and Mitchell (1998), Perkins (2000), Rosenkopf and Nerkar (2001), Katila and Ahuja (2002), Loch and Kavadias (2007), Knudsen and Levinthal (2007), and Terwiesch (2007). Our work treats innovation as a search over a landscape, with a goal of analyzing—theoretically and empirically—the underlying structure of the search space.

March (1991) and Kauffman (1993) each contribute influential models of search spaces. These models are multi-dimensional, abstract spaces. March (1991) uses the complexity of the space to introduce the distinct approaches of *exploration* (considering far flung alternatives) and *exploitation* (refinement of existing alternatives). Kauffman (1993) introduced the *NK model of rugged fitness landscapes*. This theory built from evolutionary biology has been highly influential in the academic field of management strategy, based on an analogy between the fitness of an organism and the success of an organization. See, for example, work by Levinthal (1993), Koput (1997), Rivkin and Siggelkow (2003, 2007), and Knudsen and Levinthal (2007). The NK model is flexible, and it can portray both smooth, unimodal landscapes (with an "interconnectedness" parameter, the K, of 0) and chaotic sharp-peaked landscapes (high K). An insight from this literature is that landscapes characterized by high K benefit from investments in parallel search. Sommer and Loch (2004) further investigate search strategies in different types of landscapes, comparing *selectionism* (pursuing several approaches independently) and *trial and error learning* (an incremental, local search strategy). Compared to March (1991) and Kauffman (1993), their work is more directly related to innovation as opposed to organizational problem solving more generally.

However, to the best of our knowledge, this literature of search spaces and strategies has remained theoretical, with few if any efforts to characterize landscapes empirically. One exception is Fleming and Sorenson (2004), an empirical analysis of the ruggedness of the patent space, which conceptualizes invention as search over a combinatorial space.

In our research, we focus on one of the standard modes of search studied in this literature, parallel exploration. Our contribution is to develop theory about structural elements, such as size of the opportunity space, redundancy of ideas, and clusters of similar ideas, as well as to empirically measure these elements.

Process Models of Innovation

The statistical view of innovation was first developed by Dahan and Mendelson (2001). They model creation as a series of random draws from a distribution followed by a selection from the generated ideas. This approach is analogous to models of the economics of search (e.g., Stigler 1961, Kohn and Shavell 1974, Rothschild 1974, Lippman and McCall 1976, Weitzman 1979, Morgan and Manning, 1985). Two other recent papers use the statistical view. First, Kavadias and Sommer (2009) model the idea generation process and look specifically at how process design choices relate to underlying problem structure. Second, Girotra et al. (2010) develop the idea of innovation as a search for extreme values, and model innovation as independent draws from a quality distribution. Our approach also takes a statistical perspective on the opportunity space. However, as opposed to characterizing opportunities along a single quality dimension, we also address the question of coverage of the landscape of possibilities by the search process.

3. Population Model for Size of an Opportunity Space

Our approach to studying innovation also uses a process model. We focus on the process of identifying a set of opportunities, recognizing that there can be repetition in the set. That repetition provides clues to the size of the "total population" of opportunities. To understand our model, consider opportunity identification as fishing in a lake. Each draw is a catch, with the fish released back into the lake. Sometimes the same fish will be caught again. The more frequently an individual fish is caught, the smaller the estimate of the fish population. Laplace reportedly used such a model to estimate the population of France in 1802 (Cochran 1978); the technique, called the capture-recapture method, has since been adapted to wildlife ecology (e.g., Cormack 1964, Seber 1965, Seber 1982, Amstrup et al. 2005). This type of model has also been applied to problems outside of ecology, such as estimating the size of the knowledge set in brand recall, as in Hutchinson et al. (1994).

The capture-recapture method models a sequential process in which the probability that the next idea is unique (i.e., the fish has never been caught previously) is a decreasing function of the number of ideas generated.[2] That probability decay can be represented by an exponential function. We define $p(n)$ as the probability that the next idea is unique given n ideas generated already:

$$p(n) = e^{-an} \tag{1}$$

The expected number of unique ideas out of n generated, $u(n)$, is the integral under this curve. (In using the integral we are making a continuous approximation to the—obviously discrete—number of ideas.)

$$u(n) = (1/a)(1-\mathrm{e}^{-an}) \tag{2}$$

This particular form of probability decay, the exponential form, comes from a specific underlying process, one in which there are T unique ideas total (T fish in the population) and each is equally likely to be drawn. This equally likely assumption is used in the Lincoln-Peterson method (Lincoln 1930), the standard model for estimating population size in the wildlife ecology literature. Some authors have relaxed this assumption (e.g., Sudman et al. 1988). We will also relax this assumption in Section 5.

The decay parameter and the total T are linked: $T = 1/a$. This model has only a single parameter, a, and that parameter is the inverse of the very thing we are interested in, the size of the opportunity space, i.e., an estimate of the total number of unique ideas that would result if an enormous number of ideas were generated by an unlimited number of comparable idea generators.

This capture-recapture model from wildlife ecology can be used to answer one of our key questions. Given a set of ideas generated, and given a count of the number of ideas that are

---

[2] The sequential capture metaphor embodied in this model should not be confused with sequential search in innovation, in which the identification of one opportunity benefits from knowledge gained from the identification of prior opportunities. In the capture-recapture model, sequential draws are independent of each other as in parallel search in innovation.

unique in that set, the model can be used to calculate *T*, an estimate of the size of the opportunity space.

4. Data

We report results for five different data sets, each comprised of several hundred ideas. These data sets were all generated by groups of students as part of project work they were doing for our courses on product development or innovation. The characteristics of the data sets are summarized in Table 1.

All five data sets are quite similar in structure, in that all were generated in response to a similar charge to participants and all were submitted to the same web-based tool for managing ideas (http://www.darwinator.com). Each idea in each data set was described with a title and a paragraph of text. The descriptions were not limited in length, but tended to be a few sentences. An example of an idea (from the New Ventures data set) is as follows:

> **Airplane Dating**
> "Airplane Dating" is a service that would help place singles in a specified section of an airplane where other singles have registered. A profile is created and recommended matches are sent to the subscribers.

Table 1: Characteristics of the five data sets.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Description	Ventures that could be explored and prototyped in six weeks by a team of MBA students.	Web-based product or service that could be prototyped in a one-week workshop	Physical products for college student market with retail price <$50	Physical products for college student market with retail price <$50	New technologies for use in higher-education classroom instruction
Year	2007	2009	2008	2009	2008
Sample Size	232	249	290	286	311
Population	47 executive MBA students	53 executive and full-time MBA students	58 undergraduate and graduate students in business and/or engineering	58 undergraduate and graduate students in business and/or engineering	63 undergraduate business students
Quality Metric	How valuable is this opportunity?	How appealing is this opportunity to you as a potential user?	How likely is it that pursuing this opportunity will result in a great product?	How attractive would a product addressing this opportunity be to you personally?	How do you rate this concept (Hate it / Love it)?

The students in these classes were studying innovation. They were trained in idea generation methods and many if not all intended to pursue careers closely related to innovation. Two of the data sets were generated largely by mid-career working professionals participating in a weekend executive MBA program. The alumni of these courses have an impressive track record in pursuing new ventures after graduation, often based on their class projects. (See, for example, Terrapass.com, OfficeDrop.com, DocASAP.com.) Thus, we believe these data are closer to what might be derived from industrial field studies than what might be generated in laboratory experiments with untrained subjects.

There is no overlap in the participants across the five data sets. Each individual typically contributed five ideas, but individuals worked independently. However, the ideas are not strictly independent for two reasons. The first reason is within-person dependence. The within-person effect could either be that a single person will self-censor to avoid duplication in the five ideas submitted; or the effect could be the opposite, that a single person will generate ideas that are variations on a theme. We examine both of these issues in our analysis (Sections 5 and 6). The second reason is between-person dependence related to shared experience. Our analysis *assumes* a particular generating process and attempts to estimate the size of the opportunity space it has access to. Different processes would result in different sizes. For instance, imagine that the process engaged elementary school children in generating ideas for surgical instruments. Surely this process would result in different results than one that engaged engineers, or one that engaged surgeons, for instance. The ideas generated by a process are not independent in the sense that they are generated by a group of individuals who may share some characteristics like geographic location, experience, training, age, and so forth. The ideas are only independent in the sense that the generation of idea N does not depend on an observation of ideas 1 through N-1. Indeed, these ideas can be thought of as *parallel* or *simultaneous* draws. This scenario is typical of processes that collect ideas from a large number of sources without feeding back to those sources the results of the idea collection effort.

The methods and approach in the courses in which the students were enrolled generally take a "market pull" perspective on innovation. Most of the opportunities identified by the participants are therefore articulated in terms of the problem or need to be addressed. Very few of the opportunities are driven purely by the availability of a technology.

Quality Measures

The web-based submission tool used by the subjects was also used for peer evaluation of the quality of the ideas. We used the tool to aggregate 10-20 independent judgments from participants on a 10-point scale for the quality metric indicated in Table 1. The tool does not gather judgments from the originator of an idea. It is not possible to know the "true" quality of all the ideas, which would require observing the economic value created from each idea, good and bad, from an optimal investment of development resources under all the possible market and competitive scenarios which might play out. A set of 10-20 independent subjective judgments have been shown by Girotra et al. (2010) to be internally consistent and highly correlated with purchase intent and other measures of idea quality, and we believe that these evaluations are the best practical indicator of the value of the ideas.

Similarity Measures

Similarity of ideas is a central element of our conceptual framework. For our purposes we need to measure the extent of similarity between every pair of ideas within each data set. Our measurement technique was motivated by the enormity of this task. Consider, for example, the New Products I data set comprised of 290 ideas. We would like to estimate the level of similarity between each pair of different ideas in the data set. To do this, we need to make (290 x 289)/2 = 41,905 comparisons. Figure 1 is a matrix showing the results of such estimates, with each cell in the matrix representing a pair of ideas: cell (i, j) represents the pair of idea i and idea j. The darker the cell, the more similar the pair. The figure illustrates the complexity of the estimation task. Recall that we have five data sets, so in total we actually need to make about 200,000 comparisons. One way to do this would be to present pairs of ideas to judges and ask them to rate the level of similarity. For robustness, we would want to average the judgments of multiple raters for each pair. With three raters for each pair, if each judgment took only 15 seconds, this approach would require 2,500 hours of rater effort, more than a full work year, which would be prohibitively time consuming and costly.

Instead of that pair-by-pair approach, we developed a more efficient and less tedious method for measuring similarity. In our approach, respondents look at a list of ideas—titles plus descriptions—and identify groups of similar ideas. Rao and Katz (1971) document the challenges in assessing similarity between the pairs of elements in large data sets; our approach is most similar to the category of approaches they call “picking.” Based on several pre-tests, we learned that this task is manageable for lists of up to about 85 ideas, a quantity that can be printed on three letter-size sheets of paper. (With many more ideas than that, we observed that respondents faced difficulty accurately recalling the ideas well enough to identify similar groups. That limit of 85 ideas means that respondents could not be simply given the entire list of ideas and be expected to accurately identify similar groups.) Using this method, we presented raters with three-page lists of ideas and asked them to create *groups of similar ideas*. We then asked the raters to reconsider the groups of similar ideas and identify any subsets from these groups that were *identical or essentially identical*. The exact instructions to the raters are in Appendix A.

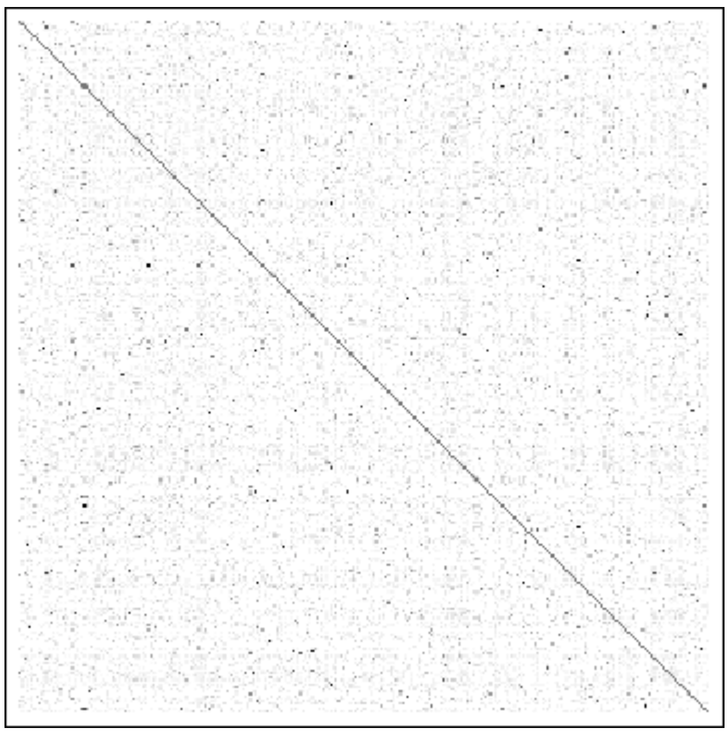


Figure 1: Similarity between pairs of ideas for the New Products I data set. The degree of similarity is represented by gray levels in each cell of a 290 by 290 matrix: cell (i, j) shows the similarity between idea i and idea j. In this data set, approximately 26% of the pairs have non-zero similarity.

We experimented with different types of questions, including coding on multiple dimensions of similarity, such as similarity of need, similarity of solution, similarity of market, similarity of function provided. However, the combinatorial complexity of the similarity coding problem is immense, and even a slight increase in the cognitive burden of the task threatened feasibility. As a result, we deliberately instructed the respondent to use his or her own notion of overall similarity in constructing groups. Other scholars reached the same conclusion about instructing participants on similarity. For example, Griffin and Hauser (1993) also leave the definition of similarity unspecified in their customer-sort procedure. More broadly, procedures for creating affinity diagrams (e.g., Kawakita 1991) call for the grouping of concepts according to the participants' own notions of similarity. Finally, Tversky (1977) advocates approaching similarity holistically, showing that empirically, similarity ratings do not correspond to underlying multi-dimensional attribute models.

We devised a method to form 40-50 different lists of about 80 ideas each from the 200-300 ideas in each data set. We formed these lists such that each pair of ideas appeared together on an average of about four different lists. These lists reflected overlapping samples of the 200-300 ideas such that most pairs of ideas appeared multiple times. The procedure for forming these lists is detailed in Appendix B.

We used university student subjects in the behavioral laboratory of one of our universities as raters. A rater was assigned a list and asked to form similarity groups. In total, we obtained 230 responses across the five data sets. The sessions were not timed and subjects were paid $10 for participating. Most subjects required 30-50 minutes to complete the similarity grouping task. As part of the protocol, we asked subjects for feedback on the task after they were finished. Many reported that the task was interesting. Some reported that the task was challenging. Very few reported that the task was overwhelming.

The net result of the similarity coding was that for each of the five data sets, we obtained a list of groupings of "similar" and "identical or essentially identical" ideas for each of 40-50 subjects and their associated lists of ideas. These similarity groupings are the raw data from which we compute various similarity measures.

With the population model (Equation 2) and the three types of data—idea descriptions, idea quality measures, and similarity ratings—we are now ready to complete the analysis addressing the key questions. Figure 2 gives a complete overview of our process: the data, the analyses (to be described subsequently), and paths to the three key questions.

List of Ideas

A GPS Bike Tracking Device
B Electronic Luggage Identifier
C Suitcase Pager
D Retractable/Detachable Heel
E Cobblestone Heel Support
F Removable Heel
G Triceps Dumbbell
H Hands-Free Forearm Strength Trainer

230 judges

Similarity Matrix

	A	B	C	D	E	F	G	H
A	-	3.71	3.33	0.00	0.00	0.00	0.00	0.00
B	3.71	-	7.50	0.27	0.00	0.34	0.00	0.00
C	3.33	7.50	-	0.27	0.00	0.00	0.00	0.00
D	0.00	0.27	0.27	-	10.00	8.00	0.00	0.00
E	0.00	0.00	0.00	10.00	-	8.33	0.00	0.00
F	0.00	0.34	0.00	8.00	8.33	-	0.00	0.00
G	0.00	0.00	0.00	0.00	0.00	0.00	-	10.00
H	0.00	0.00	0.00	0.00	0.00	0.00	10.00	-

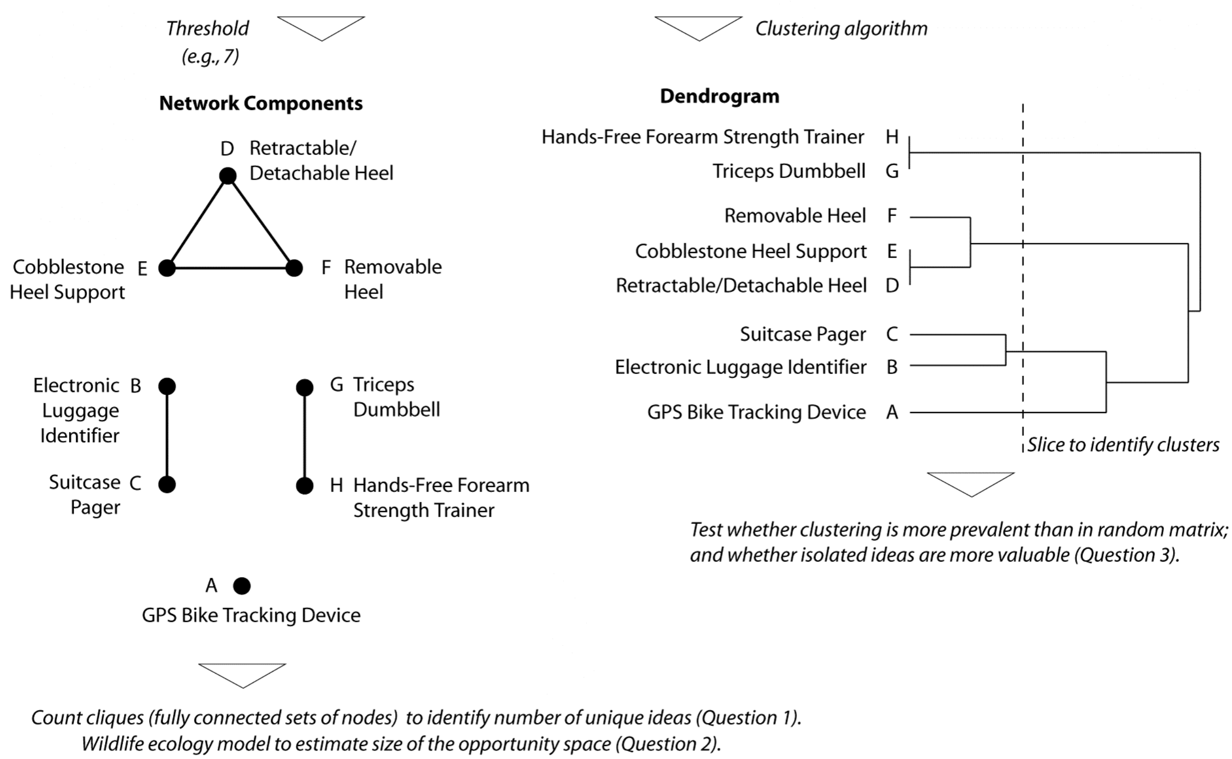


Figure 2. Analytical framework and approach. This analysis is performed for each of five independent data sets.

5. Redundancy of Ideas

The first of our key questions is about the level of redundancy in each of the data sets: how often is the exact same idea repeated? In this section, we describe how we used the raters' assessments of identical ideas to calculate redundancy. Then we show how we applied the population model (Equation 2) to estimate the size of the opportunity space, the total number of unique ideas. Finally, we address several issues related to the robustness of that estimate: confidence intervals; relaxing the equally likely assumption of the model; and controlling for the fact that each person typically generated five ideas, which adds a sequential element to what is largely a parallel search.

Determining the Number of Unique Ideas

To measure redundancy, we identify clusters of "identical" ideas within each data set. For this analysis, we use only the groupings of *identical or essentially identical ideas* provided by each rater. A pair of ideas is defined as identical when enough raters who saw the pair rate it as identical.

To ensure robustness, we apply two different thresholds. The *majority threshold* is defined as 50% of the raters on whose lists of ideas the pair appears. The *consensus threshold* is defined as 70% of the raters on whose lists the pair appears. Thus, for a pair to be coded as identical under the majority threshold, 50% or more of the raters exposed to the pair would have grouped the pair together as identical, and for the consensus threshold 70%. These are of course arbitrary cut-offs for the definition of identical, which is why we report results for two different thresholds.

In applying these thresholds, we exclude from consideration outliers, defined as any groupings of "identical" ideas that are larger than the 95th percentile of group size for the data set in question. We do this because one or two raters for each data set constructed extremely large groups of "identical" ideas. For example, one rater constructed a group of 49 ideas, all rated as "identical or essentially identical" to one another, reflecting either a disregard for instructions or a very unusual definition of *identical*. Culling these outliers is important because otherwise each of the 49x48/2 = 1176 pairs of ideas would count in the computation of the similarity metric. Thus, very large groups of identical ideas are not only implausible, but they disproportionately influence the metric.

Here we give an example of the outcome of this analysis for one data set, New Products I. Then, we summarize the results of the analyses in a table for the other data sets. There are 290 ideas in the New Products I data set. Of these, 197 are *not identical* to any other idea using the majority threshold. That is, for each of these 197 ideas, there is no other idea deemed identical to that idea by half or more of the raters. The remaining 93 ideas are clustered into the twenty-four network *components* shown in Figure 3. (In network analysis, a *component* is a group of nodes that are interconnected, at least indirectly, and that are not connected to other nodes [Scott, 2000].) There are 11 *pairs* of ideas; 4 *triples*; 4 clusters of four; and so forth. The distribution of sizes of network components for all five data sets is shown in Table 2.

The distributions presented in Table 2 show that the level of redundancy in the data sets is quite low. Even at the majority threshold, which reflects a fairly loose notion of what it means for two opportunities to be identical, most ideas are not considered identical to any other idea in four of the five sets, all but Classroom Technologies. At the consensus threshold, 85-90% of the ideas in the first four data sets are not considered identical to any other. And even in Classroom Technologies, with the most narrowly defined scope, 68% of the ideas are not considered identical to any other.

To apply our model to estimate the size of the opportunity space, i.e., the *total* number of unique ideas, we need an estimate of the number of unique ideas *within* each data set. Simply counting the number of components in the network would understate the number of unique ideas. Because of the multi-dimensionality of similarity and the latitude in the threshold, identical relationships are not fully transitive. Therefore, not all ideas in every component are identical. For example, the Backpack/Umbrella appears in the same component (seen in the upper left corner of Figure 3) as the Hands Free Coffee Sleeve, and yet clearly these are two different ideas. We use the definition of a *clique* from network analysis to count the number of unique ideas. A clique is a *fully connected* set of nodes: every node in the set is directly connected to every other node in the set (Scott 2000). If a set of ideas is truly identical, then those ideas should appear as cliques in the network.

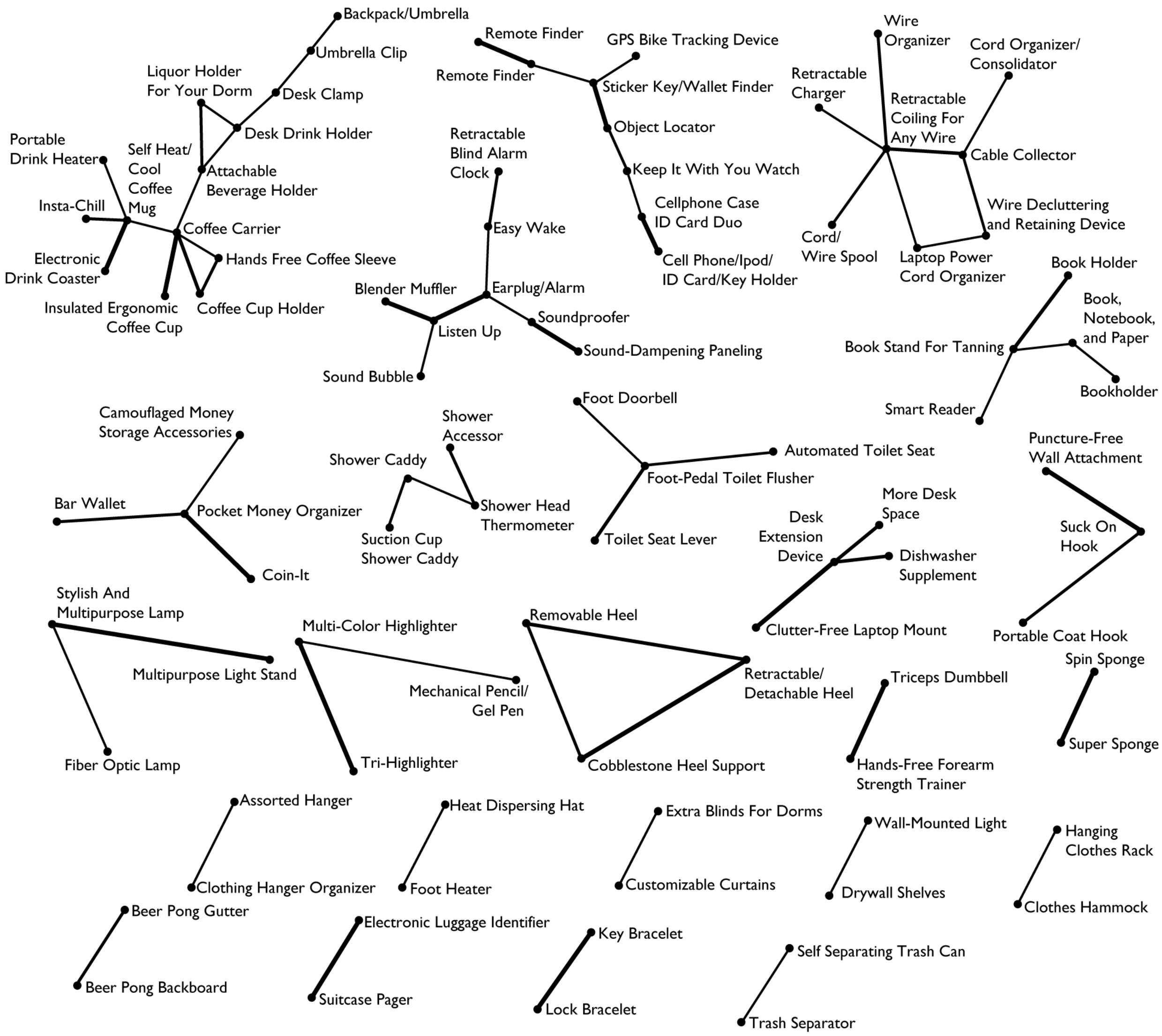


Figure 3. Clusters of identical ideas for data set New Products I based on the majority threshold for the definition of identical. The 197 singletons (i.e., ideas for which there are no identical counterparts) are not shown. The thickness of the links is proportional to the fraction of raters identifying the pair as identical. The labels are the actual titles used by the originator of the idea, and so do not always summarize the description of the actual opportunity precisely.

Table 2. Distribution of network component sizes for each data set and for two thresholds defining *identical*. The value of N is the number of ideas in components of a given size (i.e., 15 clusters of 2 is shown as N=30).

Panel A: Majority Threshold for Identical (≥50% of raters identify pair as identical)

	New Ventures		Web-Based Products		New Products I		New Products II		Classroom Technologies	
	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas
Singletons	139	60%	175	70%	197	68%	165	58%	78	25%
Pairs	30	13%	40	16%	22	8%	40	14%	6	2%
Triples	12	5%	12	5%	12	4%	27	9%	6	2%
Clusters of 4	12	5%	12	5%	16	6%	4	1%	0	0%
Clusters of 5	0	0%	10	4%	5	2%	5	2%	0	0%
Clusters>5	39	17%	0	0%	38	13%	45	16%	221	71%

Panel B: Consensus Threshold for Identical (≥70% of raters identify pair as identical)

	New Ventures		Web-Based Products		New Products I		New Products II		Classroom Technologies	
	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas	N	Fraction of Ideas
Singletons	206	89%	224	90%	247	85%	243	85%	213	68%
Pairs	20	9%	16	6%	30	10%	30	10%	32	10%
Triples	6	3%	9	4%	9	3%	9	3%	12	4%
Clusters of 4	0	0%	0	0%	4	1%	4	1%	4	1%
Clusters of 5	0	0%	0	0%	0	0%	0	0%	5	2%
Clusters>5	0	0%	0	0%	0	0%	0	0%	45	15%

We count the cliques from largest to smallest. First we find the largest clique (fully connected set of nodes), count that as a single idea, and remove it from the network. Then we identify and remove the largest clique in the remaining network, and so forth, until there are only singletons left. Each singleton naturally counts as a unique idea. We break ties by randomly selecting a largest clique.

Finding the cliques in a network is an NP-hard problem (Karp, 1972). However, the identical matrices are very sparse (i.e., most of the links are 0), so we were able to complete the computations. This approach has been used in network analysis applications such as identifying community structure (Yan and Gregory, 2009) and creating reduced forms of large networks for visualization (Six and Tollis, 2001).

The results of our count of number of unique ideas for each data set are shown in Table 3.

Table 3. Estimates of number of unique ideas for each data set based on counting cliques in the identical network, at the majority threshold and consensus threshold.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Ideas in data set (*N*)	232	249	290	286	311
Number of unique ideas (*u*) at majority threshold	191	216	252	231	216
Percent unique	82%	87%	87%	81%	69%
Number of unique ideas (*u*) at consensus threshold	220	238	271	267	271
Percent unique	95%	96%	93%	93%	87%

Applying the Model to Estimate the Size of the Opportunity Space

Using the tally of unique ideas, we can now estimate the a parameter of the population model (Equation 2) for each data set. Each data set has a size, N, and a number of unique ideas in that set, u. These two numbers, (u, N) produce an estimate of a from a numerical solution[3] to $u = (1/a)(1\text{-}e^{-aN})$. The expected total number of unique ideas is then calculated as $T = 1/a$. In Table 4, we show those values for the consensus threshold on identical. (The T values are rounded in the table.)

Figure 4 illustrates the relationship between the number of unique ideas identified and the total number of ideas generated for two of the data sets. The relationship is concave; it is increasingly difficult to identify unique ideas as the number of ideas generated increases. Different domains and generating processes would exhibit different curves.

[3] Dawkins (1991) gives an approximation to T as $u^2/(2(n\text{-}u))$. For the first four data sets, this approximation underestimates T by about 10%; for the fifth one, it underestimates by nearly 20%.

The notion of finite number of unique ideas needs to be qualified. In a real sense, the number of ideas is not finite. There is an arbitrarily large number of attributes that can be used to characterize an opportunity (e.g., focal user segment, performance level, nuances of needs addressed, etc.). Within our analytical framework, the identical threshold defines a level of resolution beyond which two ideas are categorized as the same idea. This qualifies the definition of *T* as the total number of ideas that are distinct enough from one another to exceed that threshold. With that qualification, we can reasonably consider the size of the opportunity space to be finite.

Table 4. Estimates of total number of unique ideas, *T*, in each opportunity space based on values for *N* and *u* for each data set.

	New Ventures	**Web-Based Products**	**New Products I**	**New Products II**	**Classroom Technologies**
Ideas in data set (*N*)	232	249	290	286	311
Number of unique ideas (*u*) at consensus threshold	220	238	271	267	271
Parameter (*a*) estimate	0.000462	0.000366	0.000473	0.000486	0.000907
Estimate of *T*, total number of unique ideas	2165	2735	2115	2056	1103
Lower bound for *T* (2.5th percentile)	1205	1493	1333	1299	806
Upper bound for *T* (97.5th percentile)	3704	4762	3333	3226	1493

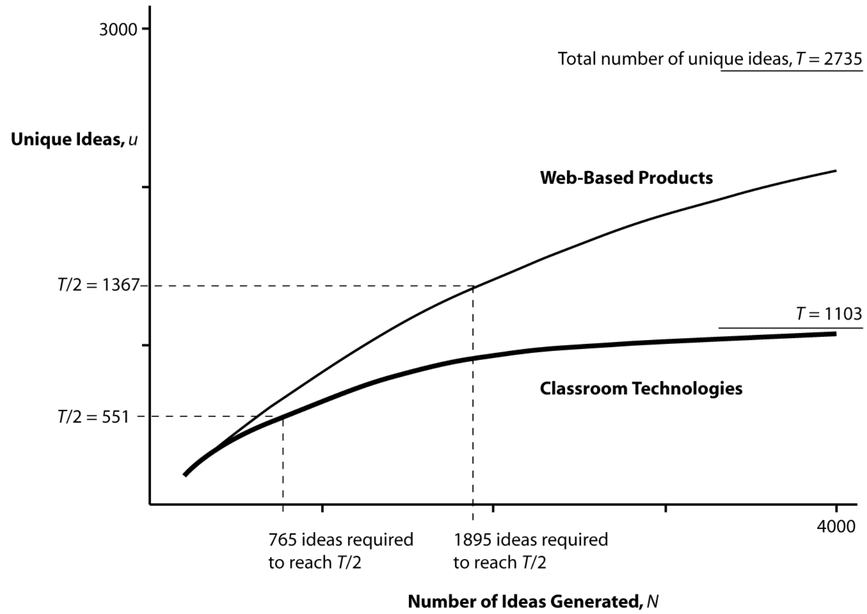


Figure 4: Number of unique ideas, u, expected for a given number of ideas generated, N. Two domains are shown, Web-Based Products and Classroom Technologies.

Confidence Intervals

Using our model we have derived point estimates of the total number of unique ideas, T, for each data set. Our model for the probability that the next idea is unique (Equation 1), dictates a stochastic process for the number of unique ideas in any set. Using that uncertainty, we can numerically approximate confidence intervals around our estimates of T. The details of how we do this are explained in Appendix C.

The results for the 95% confidence intervals are shown in the last two rows of Table 4, rounded to the nearest whole number. The confidence intervals are wide, but appropriately so: they reflect the level of uncertainty in the process.

We test whether the estimated sizes of the opportunity spaces are statistically significantly different. We find that the sizes of the first four opportunity spaces are not statistically different from one another, and the first four are all statistically significantly greater than Classroom Technologies (with three of the four at the 0.05 significance level and Web-Based Products at the 0.01 level). Details are in Appendix D.

This test confirms the intuitive notion that the Classroom Technologies space is a smaller or narrower space. The innovation charge for the Classroom Technologies domain cued both a “how” (“technology”) and a setting (higher education classroom), so there is a base level of similarity across every single idea. In contrast, the innovation charges for the other domains were more abstract, soliciting ideas for general products and ventures.

Relaxing the Equally Likely Assumption

Now we return to one of the fundamental assumptions in landscape size estimation: what if the ideas are not equally likely? A logical replacement for the equally-likely assumption is an empirical distribution based on the observed relative frequency of the unique ideas in each data set. To construct that relative frequency distribution, we use the clique sizes for each of the unique ideas identified in each data set. In considering different levels of T (total number of unique ideas), we stretch (or shrink) the distribution accordingly. Using a grid search, we find the T that gives the best match with the observed data for each set. Matches are determined by repeatedly simulating N draws from a population of size T according the relative frequency distribution of clique sizes, and looking for the value of T that results in $u(N)$ unique ideas (e.g., 271 for New Products I at the consensus threshold). The estimates of T based on this approach are shown in Table 5, along with the original estimates based on the equally likely assumption. The estimates of T do not change much with this analysis. In every case, accounting for the non-uniform distribution raises the estimate somewhat.

Table 5: Estimates of the total number of unique ideas, T, based on empirical relative frequency of ideas. These estimates use the consensus threshold for identical, 1000 simulation trials, and a grid search interval of 15.

	New Ventures	**Web-Based Products**	**New Products I**	**New Products II**	**Classroom Technologies**
T assuming each idea equally likely	2165	2735	2115	2056	1103
T assuming empirical frequency distribution	2268	2839	2334	2205	1192

Robustness to Multiple Ideas per Person

Our model of unique idea generation captured in Equation 2 is based on a process in which each idea is a draw from a pool of T equally likely unique ideas. We have already examined relaxing the equally likely assumption. Now we examine another issue in light of our data collection process, that of multiple ideas per person.

In our idea generation assignments, each student was asked to contribute five ideas. Conceptually, this can raise an issue for our data analysis. Self-censoring occurs such that a single person is highly unlikely to submit two redundant ideas. Could this explain why the level of redundancy that we find in the data sets is so low?

We examine this possibility by simulating an idea generation process in which each person generates enough ideas to have five unique ideas. The predicted number of unique ideas from Equation 2 based on the larger N that would result from this process is virtually identical to our reported results. Further details from the simulation can be found in Appendix E. This result makes sense, because the probability of encountering a redundant idea in just five draws is very low; thus the effect of censoring does not influence the main result very much.

6. Clusters of Similar Ideas

In the previous section we analyzed redundancy, the strict repetition of ideas. Now we turn our attention to a looser sense of repetition, similarity among ideas. The analysis we did for strict redundancy produced an estimate of the total number of unique ideas. We do the same analysis at a higher level of abstraction, counting the number of idea clusters in each data set and using the population model to estimate the total number of clusters in the landscape. We also show that clustering is a statistically significant feature of the landscape as compared to a random benchmark.

Computing the Similarity of Each Pair of Opportunities

Recall that we asked each of the 230 raters to group separately the identical ideas and the similar ideas. To construct clusters of similar opportunities for this analysis, we compute a similarity measure for each pair of opportunities within a data set. This similarity measure is a weighted function of the identical groupings and the similar groupings of each respondent, averaged over the respondents who had the pair on their list.

Weighted similarity is a metric ranging from 0 to 10, defined as the average over all raters of the maximum of

- 10, if the rater identified the pair as identical, and
- 15/list-length, where list-length is the length of the shortest list in which a rater included the pair.

As in our analysis of identical ratings, we exclude the top five percent longest identical lists from these calculations.

The extreme value of 10 occurs when all raters identify a pair as identical. The logic of the second term in constructing the metric (i.e., 15/list-length) is that all else equal, the longer the list of similar ideas, the more general the categorization of ideas. In previous work, respondents have been given a specified list length or a maximum list length (Rao and Katz, 1971, Methods 4 and 5). In our method, the respondent has more control over the definition of similarity.

To illustrate the logic of controlling for list length, consider dorm room storage. Lists of broad dorm room storage solutions will be longer than lists of easy-to-hang shelves. If the rater formed a group of just two similar ideas, then the similarity score for that rater and pair would be 15/2 = 7.5. If that pair of ideas were included in a group with one other idea, then the similarity score would be 15/3=5. We used the value of 15 so that the highest score a pair of ideas could receive from a similarity ranking, absent an identical ranking, was 7.5.[4] This is a scaling factor that allows both groups of identical ideas and groups of similar ideas to be used to compute a single similarity metric. Our preliminary investigations revealed that our results are not highly sensitive to this scaling factor.[5] When averaged across all raters, the weighted similarity score exhibits a relatively smooth unimodal distribution, skewed towards 0, and with a thin tail stretching to the maximum value of 10.

---

[4] Note that raters were instructed that ideas can appear on multiple lists. The similarity score for a pair of ideas comes from the shortest list on which a rater included the pair.

[5] Table 8 refers to more details of this sensitivity analysis.

The result of this computation is a similarity matrix for each domain, of which Figure 1 an example.

To evaluate how consistently different raters perceived the pairs of ideas, we calculated the variance in ratings for each pair. For example, if a pair appeared on five lists, and was rated identical (10) by two raters, similar to one idea by one rater (15/2 = 7.5), similar to two ideas by another rater (15/3 = 5), and not similar by the fifth rater, the variance in rating for that pair is the variance of (10, 10, 7.5, 5, 0) = 17.5. In each data set, we averaged the variances across all pairs of ideas. The results are shown in Table 6, and indicate an overall high level of agreement across raters.

Table 6: Variance in similarity ratings across raters for each data set.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Average inter-rater variance across all pairs of ideas	0.58	0.39	0.33	0.39	1.3

Clustering Similar Opportunities

Once we built the similarity matrices for each data set, we used them to find clusters of similar ideas. To identify clusters, we used a hierarchical clustering analysis, implemented in Mathematica. The clustering analysis iteratively groups the closest ideas, and then sets of ideas, using the average proximity (in our case the similarity score) of items in sets. The output of that analysis is a dendrogram, a tree, which displays the most similar ideas together and indicates by branches how similar the ideas are. As an example, a portion of a dendrogram for the New Products I data set is shown in Figure 5. Uses of this clustering technique are described in Punj and Stewart (1983), Girvan and Newman (2002), and Gulbahce and Lehmann (2008).

We then apply the ordering of the opportunities in the dendrogram to the order of the rows and columns in the similarity matrix, which results in clusters of opportunities appearing visually as blocks along the diagonal of the matrix as shown in Figure 6. We have labeled some of these blocks according to the opportunities they contain.

We observe that the themes that characterize the clusters in the two New Products data sets are, as one would expect, quite similar. These data sets were created by successive offerings of the same course using the same innovation charter. Both have clusters of ideas around general areas like dirty dishes, bathrooms, food and beverage, alarm clocks, school supplies, and dorm room storage. And both have clusters of ideas around more specific needs like transporting small items such as keys and IDs, managing messes of cords and wires, and locating lost objects. For many of these clusters, not only are the idea groupings similar across the two data sets, but the relative proportions of the ideas in the data set are too. For example, both have about 5% of the ideas related to the bathroom, about 10-15% related to food and beverages, and about 2-3% related to transporting small items.

Despite substantial overlap in the clusters, there are still differences in the data sets. For example, New Products I contains many ideas related to bicycling and New Products II contains almost none. These cross-set observations echo our findings that we should expect both similarity and uniqueness in idea generation.

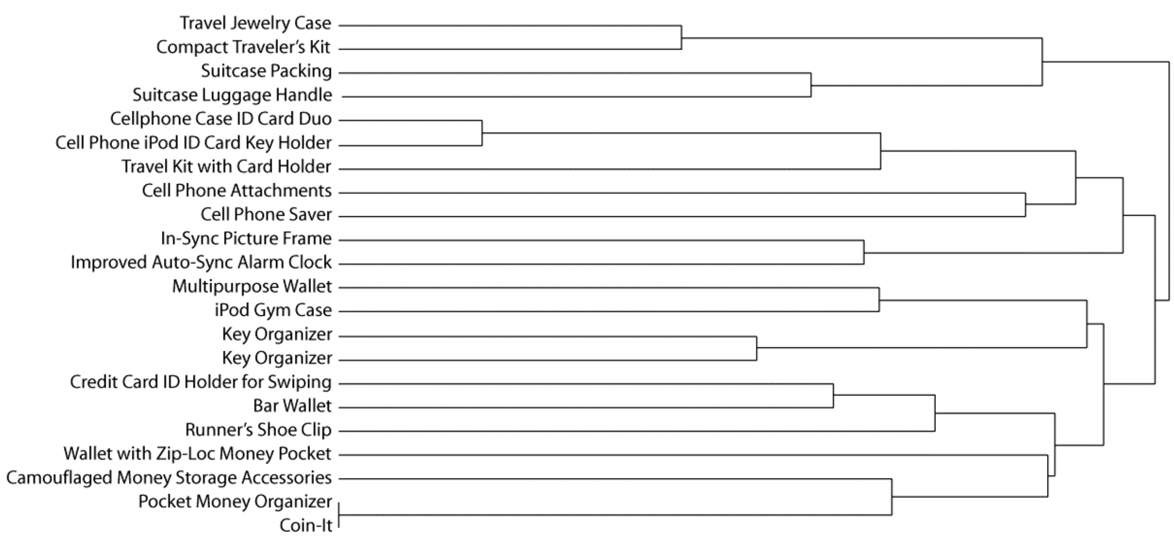
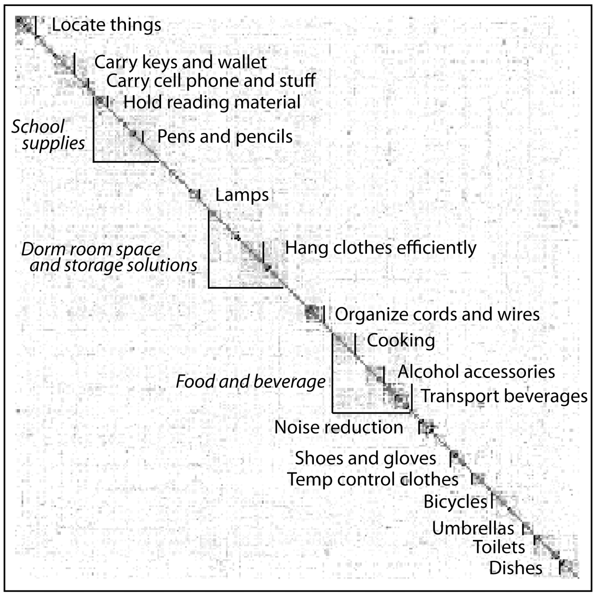


Figure 5: A portion of the dendrogram for the New Products I data set.



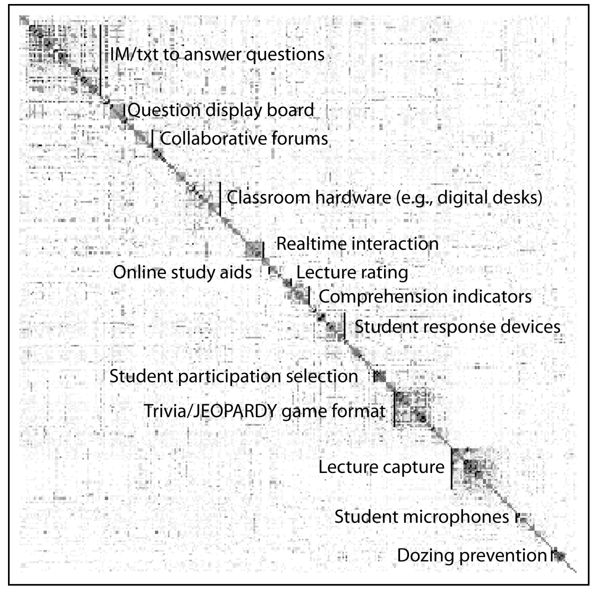


Figure 6: Reordered matrices of opportunities for the New Products I dataset (left) and Classroom Technologies showing labeled blocks along the diagonal.

Dendrogram Slicing and Estimating the Total Number of Clusters in the Landscape

By making a vertical "slice" through the dendrogram, we identify the different clusters (or branches) of the tree. If the cut is made very near the leaves of the tree (the left side of the tree in Figure 5), then the number of clusters will be high, approximating the number of unique ideas counted using cliques. If the vertical cut is made near the root of the tree, then the tree will be divided into a few, large clusters. The location of the cut determines the level of abstraction at which clusters are defined, and is a decision variable in the analysis to be performed.

For our data sets, we report on clusters at two different levels of abstraction, 1/5 of the distance from the root to the leaves and 1/10 of that distance. Slicing a dendrogram at the 1/5 distance yields clusters defined by a fairly specific shared need. For example, from Figure 5, a slice at the 1/5 mark clusters together Travel Jewelry Case and Compact Traveler's Kit—both solutions for carrying specific items while traveling—but separates those two from a clustered pair of other travel-related ideas, Suitcase Packing and Suitcase/Luggage Handle—which relate more to the logistics of the travel bags themselves. This level of abstraction is somewhere between the very strict redundancy measures used to count number of unique ideas and looser category levels.

Slicing a dendrogram at the 1/10 distance yields more general categories or clusters. At this level, the clusters relate to a more general category (e.g., travel) or purpose (e.g., carrying small items). Because the slice distance is a decision variable in the analysis, any conclusions about clustering must be accompanied by a specification of the slice distance used to define that clustering. In Figure 6, most labels correspond to selected clusters at the 1/10-slice level, chosen for their notable visual presence in the matrix. The italicized labels for New Products I in that figure show supersets at the 1/50-slice level.

Table 7 shows the number of clusters in the data sets at these two levels of abstraction, and includes an estimate of T, the total number of clusters in the landscape. This value of T is estimated from the number of ideas generated, N, considering the number of clusters as u, and adjusting for the empirical relative frequency of the ideas as explained in Section 5. In other words, in this analysis multiple ideas appearing in a cluster correspond to repeated "capture" of that cluster. At the shared-need level (1/5), there is still quite a bit of undiscovered territory ($T - u$) in all the data sets, but at the category level (1/10), most of the categories have been identified in all the data sets, and especially in Classroom Technologies.

Table 7: Estimated total number of clusters in the landscape for the five domains at two different levels of abstraction, the level of shared need (1/5 slice distance), and the level of shared category (1/10 slice distance).

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Ideas in data set (*N*)	232	249	290	286	311
Shared-need slice level (1/5)					
Number of clusters in data set (*u*)	110	133	147	147	99
Estimate of total number of clusters in the landscape (*T*)	158	201	225	228	116
Shared-category slice level (1/10)					
Number of clusters in data set (*u*)	69	84	88	98	62
Estimate of total number of clusters in the landscape *(T)*	82	100	103	112	69

Clustering as a Significant Feature of the Landscape

There are clearly clusters in the data as shown in the dendrogram. However, there would be clusters in random data as well. To support the idea that clusters represent real underlying themes in the idea generation effort, we show evidence that the opportunities are more tightly clustered than one would expect from a random sample. We address this question by comparing the clustering of the opportunities from our data sets with that which we observe on average in 50 randomly generated similarity matrices. The random matrices are generated to have the exact same cell values as the similarity matrix for a data set, but in a randomized order.[6] We can then compare the clustering in these randomly generated matrices with the clustering observed for our data sets. More formally stated, we test the hypothesis that opportunities in the data sets are more clustered than random opportunities with the same degree of similarity. Table 8 reports the results of this hypothesis test in the form of a T-test.

We find strong support for the clustering hypothesis. In every case, the number of clusters in the data sets is lower than the average number of clusters in the random benchmarks. Therefore, we conclude that opportunities generated in practice are clustered, as opposed to randomly or uniformly distributed. This is especially true at the category level. This finding suggests that there are significant underlying themes driving idea generation, and that the clustering approach can usefully identify those themes.

[6] We also analyzed random benchmarks that treat within-person and between-people pairs separately. These benchmarks replicate the actual number of individuals and number of ideas per individual in each data set, and they pull separately from the within-person similarity values and between-people similarity values. See Appendix F.

Table 8: Comparison of number of clusters in actual similarity matrices compared with the clustering in random matrices. The T-statistic compares the observed number of clusters to the distribution of clusters observed for 50 randomly generated matrices with the same values of inter-idea similarity as found in the data.[7]

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Shared-need slice level (1/5)					
Number of clusters, actual	110	133	147	147	99
Number of clusters, random (average over 50 samples)	121	139	160	152	126
T-statistic	4.24***	2.11**	3.75***	1.73*	9.29***
Shared-category slice level (1/10)					
Number of clusters, actual	69	84	88	98	62
Number of clusters, random (average over 50 samples)	82	95	109	105	76
T-statistic	4.49***	3.91***	8.07***	2.54**	5.76***

*** p<0.01, **p<0.05, *p<0.10 two-tailed tests

7. Quality and Similarity

In this section we address the third key question of the paper: are unique ideas more valuable? On the one hand, the existence of many similar ideas suggests that an idea is not truly novel, perhaps even obvious, and therefore not especially valuable. On the other hand, the existence of similar ideas might indicate that the idea addresses a widely held need, suggestive of market acceptance of the innovation. Thus, we have conflicting theoretical bases for hypothesizing the direction of a relationship between value and similarity. To capture the alternative effects, we pose the *Uniqueness Hypothesis* that the estimated value of an idea *decreases* with the number of similar ideas; and the *Popularity Hypothesis* that the estimated value of an idea *increases* with the number of similar ideas. To test these hypotheses, we regress the estimated value of each opportunity against the size of the dendrogram cluster in which that opportunity resides.

[7] In Appendix G, we show results of a sensitivity analysis to the similarity metric. First, we examine sensitivity to the scaling factor (15 in the base case). Second, we examine sensitivity to the functional form of the metric: we rerun the analysis of Table 8 for a similarity metric in which we do not adjust for list length. In those cases, we treat the similar lists like the identical lists: all pairs that appear together on a list get a fixed similarity value. As in our identical analysis, we omit the longest 5% of lists from this calculation

The dependent variable for this regression is the rating given by a specific rater to a specific opportunity. This dependent variable is an integer value between 1 and 10. Although strictly speaking, the bounds on the dependent variable violate the assumptions of ordinary least squares regression, in practice, the dependent variable rarely takes on values of 1 or 10, and exhibits a unimodal distribution well within the bounds of 1 and 10.

We control for the identity of each rater with a dummy variable, because the raters typically use different parts of the 1-10 quality rating scale.

For the cluster-size variable, we show results for two dendrogram slice levels, the shared-need level (1/5 slice) and the shared-category level (1/10 slice). The results are similar for a stricter definition of similarity (e.g.,1/2 slice). The summary statistics for the variables are in Table 9 and the results of the regressions are in Table 10. Recall that the questions used to assess the value of ideas were somewhat different for each data set, although Girotra et al. (2010) show that the responses to these questions are highly correlated.

Five out of the ten of these tests show support for the Popularity Hypothesis, that value is increasing in the number of similar ideas related to the need or in the category. None of the remaining ones show significant support for the Uniqueness Hypothesis, that value is decreasing in the number of similar ideas. In four of the five data sets, the cluster sizes produced by at least one of the slice levels (1/5 or 1/10) is a significant, positive predictor of value.[8] Even though not extremely consistent or compelling, the best single model of these data would be that value is increasing in similarity. Thus, we can reject the Uniqueness Hypothesis. There is no support for the theory that more novel ideas are considered more valuable than those that are similar to others. We consider the implications of these results in the discussion section.

[8] We also tested non-linear models (e.g., including the square of the cluster size). These models do not consistently offer more explanatory power.

Table 9: Summary statistics for variables.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Question for assessing value	How valuable is this opportunity?	How appealing is this opportunity to you as a potential user?	How likely is it that pursuing this opportunity will result in a great product?	How attractive would a product addressing this opportunity be to you personally?	How do you rate this concept (Hate it / Love it)?
Mean - Estimated Value	5.24	4.64	5.36	4.37	5.62
S.D. – Estimated Value	2.27	2.70	2.35	2.74	2.30
Mean –Cluster size (shared need, 1/5 level)	2.86	2.52	2.95	2.79	6.08
S.D. – Cluster size (shared need, 1/5 level)	1.60	1.52	2.01	1.76	5.83
Mean - Cluster size (shared category, 1/10 level)	4.68	4.30	5.16	4.66	10.07
S.D. – Cluster size (shared category, 1/10 level)	2.44	2.82	3.28	3.13	7.40
Pearson Correlation Value | Cluster size (shared need, 1/5)	0.023	-0.018	0.042	.055	0.024
Pearson Correlation Value | Cluster size (shared category, 1/10)	0.030	-0.010	0.021	.064	0.039

Table 10: Results of regression of the value rating of an opportunity as a function of cluster size, using similarity dendrogram slice levels of 1/5 and 1/10. Specific Rater IDs are included as controls. T-statistics are in brackets.

	New Ventures		Web-Based Products		New Products I		New Products II		Classroom Technologies	
Constant[9]	3.98*** [19.13]	3.95*** [18.95]	6.16*** [7.88]	6.14*** [7.85]	3.29*** [11.06]	3.36*** [11.31]	2.51*** [5.89]	2.46*** [5.80]	7.83*** [26.23]	7.77*** [25.96]
Cluster size (shared need, 1/5)	0.025 [1.36]		-0.032 [-1.31]		0.048*** [2.81]		0.064*** [2.99]		0.007 [1.29]	
Cluster size (shared category, 1/10)		0.022* [1.82]		-0.011 [-0.81]		0.015 [1.410]		0.054*** [4.47]		0.013*** [2.86]
(+ controls for raters)										
N	4626	4626	4477	4477	3366	3366	3801	3801	4189	4189
Adj. R^2	0.22	0.22	0.17	0.17	0.30	0.30	0.28	0.29	0.17	0.17
F Statistic	28.36	28.40	17.44	17.42	24.23	24.10	26.17	26.43	10.72	10.82

*** p<0.01, **p<0.05, *p<0.10 two-tailed tests

[9] The constant reported for each model is determined by which of the Rater IDs serves as the baseline for the rater dummy, and so should not be interpreted as a meaningful difference across the data sets relative to the hypotheses.

8. Discussion

To understand and characterize opportunity spaces, we tackled three main questions in this paper: (1) When a large number of independent efforts to generate ideas are conducted in parallel, how likely are the resulting ideas to be redundant? (2) Using redundancy as a clue, how vast are the opportunity spaces we study? (3) Are the less similar ideas more valuable than ideas that are relatively common? The answer to the first question is that while there is clearly some redundancy in the ideas generated by aggregating parallel efforts, this redundancy is quite small in absolute terms, even for very narrowly defined domains. For the second question, we find that the estimated total number of unique ideas is about one-thousand for one narrowly defined domain and greater than two-thousand for the other more broadly defined domains. On the third question, we find that ideas that are more distinct from other ideas are not generally considered more valuable.

In addition to answering these key questions, we have developed methods for measuring similarity, defining unique ideas, estimating the sizes of opportunity spaces, and identifying clusters of ideas. These methods have proven useful scientifically, and offer promise in practice as well.

Managerial Implications

In our five data sets of ideas, there is very little redundancy. Of course we cannot extrapolate that result to all innovation efforts and claim that there will never be much redundancy. However, the results from our data sets do demonstrate the remarkable breadth of ideas that can be produced by parallel idea generation. Organizations have some control over the breadth of ideas produced by setting the scope of the innovation effort and by involving a diverse group of people. With landscape sizes comparable to our data sets, organizations can generate hundreds of opportunities and most will be unique.

The capture-recapture model offers promise for managing the idea generation effort. Examining an initial set of ideas for redundancy gives a clue to how vast the opportunity space is, as defined by the stated innovation charge and the idea generating process. Table 11 shows estimates of total number of unique ideas (T) for different numbers of ideas generated (N) and the fraction of those that are unique (f). For example, if only 95% of the first 100 ideas are unique, the estimate of the total is 966. In this scenario, the team would probably benefit from substantial further investment in idea generation, very little of which would be wasted effort. We note that this table uses the simplest assumptions: the ideas in the opportunity space are generated independently, each with equal probability. However, our estimates using the empirical distribution of ideas showed that the equally likely model underestimates the total number of unique ideas (*T*).

Table 11: Estimate of the total number of unique ideas (*T*) for a given number of ideas generated (*N*) and the fraction of those ideas that are unique (*f*).

	F						
N	**65%**	**70%**	**75%**	**80%**	**85%**	**90%**	**95%**
50	54	66	83	108	150	233	483
100	107	131	165	215	299	466	966
150	161	197	248	323	449	699	1450
200	214	263	330	431	598	932	1933
250	268	328	413	539	748	1165	2416
300	321	394	495	646	897	1398	2899
350	375	460	578	754	1047	1631	3382
400	428	525	660	862	1196	1864	3866

In posing the paper's key questions, we noted that the level of redundancy informs the decision about how much to invest in parallel search. Dahan and Mendelson (2001) focus directly on that question in a context in which each concept is unique. Their estimates are therefore an upper bound for the number of ideas to generate when we allow for the redundancy that is likely in industrial practice.

While redundancy in our data is low, we did found strong evidence of clustering. A description of that clustering may be useful in practice: the dendrogram clustering and the implied cluster labels (as shown in Figure 6) organize several hundred ideas into a few dozen themes. Clustering has implications for the design of an innovation tournament (Terwiesch and Ulrich 2009). If each idea has to be evaluated in isolation, efficiency must be favored over depth in the evaluations. However, if clusters rather than individual ideas can be evaluated, the depth of analysis can be increased.

Our clustering analysis was originally motivated by scientific inquiry. However, the resulting dendrograms and ordered matrices provide a valuable window into the innovation process. The clusters reveal where most exploratory effort is being directed. The degree to which clusters align with the innovator's strategic intent may provide an effective diagnosis of problems in the opportunity identification processes of the innovator. More broadly, the set of ideas taken as a whole may contain information. The set structure speaks to the relative salience of different needs. We have used the clustering analysis with a major automobile manufacturer to explore the future of "electric mobility." The innovation charter was loosely defined in the sense that any ideas related to the future of transportation and innovative technologies were entertained. This laxity was daunting to the company at first. However, the clustering analysis revealed themes, making the structure of the opportunity space come into focus. The clusters then served as a useful tool in framing the evaluation phase.

We observe that when generating ideas with practicing professionals, there appears to be an instinctive positive response to unique ideas. This response appears to be even more pronounced with novice innovators, who often dismiss a cluster of ideas because the similarity of those ideas mean that they do not seem sufficiently novel. Our data show that this reaction may be at odds with the evidence that unique ideas are not systematically valued more highly than ideas that are similar to others. This result implies that managers should pay closer attention to the message that repetition in idea generation may be signaling a strongly felt need.

Limitations

There are four main limitations to this research. First, these data are derived from a classroom setting. While about half of our subjects were mid-career professionals and experienced innovators, they were still working within an educational setting. It would be interesting to do a similar analysis of a data set arising naturally from commercial activity, as one might find in the development organization of a consumer products company. Of course our estimates of landscape size pertain specifically to the data sets we collected. Just as the ideas themselves depend on who is generating the ideas, so does the landscape size.

Second, the quality measure for our opportunities is a subjective peer evaluation. It is possible that this measure is poorly correlated with the expected value of the eventual commercial success of an opportunity if pursued. However, it is of course practically impossible to get profit outcomes for hundreds of opportunities, most of which do not warrant investment. Furthermore, even a profit outcome would be just a particular realization of a stochastic process dependent in part on exogenous factors. Prior research shows that these peer evaluations are highly correlated with purchase intent, which is reflective of one of the main drivers of eventual success—market acceptance.

Third, the similarity rating task is challenging to execute perfectly. One of the issues with our approach is that a pair of ideas might be judged more or less similar based on the other ideas with which they appear. Indeed, Ratneshwar et al. (2001) show that similarity is somewhat context-dependent.

Fourth, the innovation challenges from which our data are derived were fundamentally needs-driven endeavors. The participants possessed relatively general capabilities as entrepreneurs and product designers and were seeking out unmet market needs. While we believe that most successful innovation is market-driven, we would expect different patterns of similarity and quality for opportunities that were fundamentally technology or solution driven.

Future Work

The patterns we observe in large samples of innovation opportunities are the result of both the nature of the landscape and the nature of the search process. To what extent can the search process be managed to achieve different results? Hoffman et al. (forthcoming) suggest that it is certain customers, ones with an "emergent nature" that should be tapped by idea generation. Are there strategies that improve the idea generation performance of non-emergent customers? For example, do some heuristics for idea generation result in less clustered outcomes? Dahl and Moreau (2002) describe the positive effect of far analogies on creativity and idea value. Would innovators prompted with this knowledge produce less clustered ideas? Toubia (2006) examines how incentive structure affects creative output, another approach to managing the process.

We have only begun to probe the phenomenon of clustering. These questions seem promising for further exploration:

- How do the patterns of opportunities compare to the patterns of successful commercial innovations? What do differences between the patterns of opportunities and the patterns of existing successful products reveal? We are struck in our project-based courses by how some of the same opportunities have arisen for many years (e.g., better wire and cord management). Do these recurring gaps reveal technological limits (i.e., a very hard problem for which no good solutions have yet been developed)?

- The relationship between similarity and value is, if anything, positive. This result is consistent with there being a common driver of quality and clustering, an underlying interest or attraction from the idea generating group.[10] Further exploration of these potential underlying factors would be interesting.

- Erat and Krishnan (2010) develop a model that shows how clustering can be a consequence of a group of innovators all trying to propose the best idea. To what extent do incentives and competition drive the clustering, either at the level of individual innovators or possibly at the level of innovating firms?

- Are patterns in the opportunity landscape fractal in nature? That is, would we observe similar patterns of redundancy and clustering when examining innovation opportunities at very different levels of abstraction? These levels might extend from the level of identifying potential new businesses down to the level of identifying potential new design details on individual products.

References

Amstrup, S. C., T. L. McDonald, and B. F. J. Manly (eds.). 2005. *Handbook of Capture-Recapture Analysis*. Princeton University Press: Princeton, New Jersey, 296 ff.

Baron, R.A. and M.D. Ensley. 2006. Opportunity Recognition as the Detection of Meaningful Patterns: Evidence from Comparisons of Novice and Experienced Entrepreneurs. *Management Science*, 52(9): 1331-1344.

Cochran, W. G. 1978. Laplace's ratio estimators. Pages 3-10 in H. A. David (Ed.). *Contributions to survey sampling and applied statis*tics. Academic Press, New York.

Cormack, R. M. 1964. Estimates of survival from the sightings of marked animals. *Biometrika* 51:429-438.

Dahan, E. and H. Mendelson. 2001. An Extreme Value Model of Concept Testing. *Management Science* 47(1): 102-116.

Dahl, D.W. and C.P. Moreau. 2002. The Influence and Value of Analogical Thinking During New Product Ideation. *Journal of Marketing Research*. 39: 47-60.

Dawkins, B. 1991. Siobhan's Problem: The Coupon Collector Revisited. *The American Statistician*. 45(1): 76-82.

Diehl, M. and W. Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*. 53(3): 497-509.

[10] We thank the Associate Editor for this suggestion.

Erat, S. and V. Krishnan. 2010. Managing Delegated Search Over Design Spaces. SSRN working paper: http://ssrn.com/abstract=1567599.

Fleming, L. and O. Sorenson. 2004. Science as a Map in Technology Search. *Strategic Management Journal*. 25: 909-928.

Fleming, L., S. Mingo, and D. Chen. 2007. Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly*, 52: 443-475.

Fleming, L. and S. Mingo. 2007. Creativity in New Product Development: An Evolutionary Integration. Chapter 5 in C. Loch and S. Kavadias (Eds.). *Handbook of New Product Development Management*.

Girotra K., C. Terwiesch, and K.T. Ulrich. Forthcoming 2010. Idea Generation and the Quality of the Best Idea. *Management Science*.

Girvan, M. and M. E. J. Newman. 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA 99*, 7821–7826.

Griffin, A. and J.R. Hauser. 1993. The Voice of the Customer. *Marketing Science*. 12(1): 1-27.

Gulbahce, N. and S. Lehmann. 2008. The art of community detection. *Bioessays* 30(10): 934 – 938.

Hays, W. L. and R. L. Winkler. 1971. *Statistics: Probability, Inference, and Decision*. Holt, Rinehart and Winston, Inc.

Hoffman, D.L., P.K. Kopalle, and T.P. Novak. Forthcoming. The "Right" Consumers for Better Concepts: Identifying and Using Consumers High in Emergent Nature to Further Develop New Product Concepts. *Journal of Marketing Research.*

Hutchinson, J.W., K. Raman, and M. K. Mantrala. 1994. Finding Choice Alternatives in Memory: Probability Models of Brand Name Recall. *Journal of Marketing Research*. 31:441-461.

Kallenberg, O. *Foundations of Modern Probability*. New York: Springer-Verlag, 1997.

Karp, R. M. 1972. Reducibility Among Combinatorial Problems, in R. E. Miller and J. W. Thatcher (Eds.) *Complexity of Computer Computations*. New York: Plenum. pp. 85–103.

Katila, R. and G. Ahuja. 2002. Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. *Academy of Management Journal*, 45: 1183-1194.

Kauffman, S.A. 1993. *The Origins of Order*. Oxford University Press, New York.

Kavadias, S. and S. C. Sommer. 2009. The Effects of Problem Structure and Team Expertise on Brainstorming Effectiveness. *Management Science*. 55(12): 1899-1913.

Kawakita, J. 1991. *The Original KJ Method*, Kawakita Research Institute. Tokyo.

Knudsen, T. and D.A. Levinthal. 2007. Two Faces of Search: Alternative Generation and Alternative Evaluation. *Management Science*. 18: 39-54.

Kohn, M.G. and S. Shavell, 1974. The Theory of Search. *Journal of Economic Theory* 9(2): 93-123.

Koput, K.W. 1997. A Chaotic Model of Innovative Search: Some Answers, Many Questions. *Organization Science*, 8: 528-542.

Levinthal, D. 1997. Adaptation on Rugged Landscapes. *Management Science,* 43: 934–950.

Lincoln, F. C. 1930. Calculating waterfowl abundance on the basis of banding returns. US Department of Agriculture Circular, No. 118.

Lippman, S. A. and J. J. McCall. 1976. The Economics of Job Search: A Survey. *Economic Inquiry* 14(2): 155-189.

Loch, C.H. and S. Kavadias. 2007. Managing New Product Development: An Evolutionary Framework. Chapter 1 in C. H. Loch and S. Kavadias (Eds.) *Handbook of Research in New Product Development Management*, Butterworth/Heinemann (Elsevier), Oxford, UK.

March, J. 1991. Exploration and Exploitation in Organizational Learning. *Organization Science*, 2: 71-87.

March, J. and H. Simon. 1958. *Organizations*. Blackwell: Cambridge, MA.

Martin, X. and W. Mitchell. 1998. The Influence of Local Search and Performance Heuristics on New Design Introduction in a New Product Market. *Research Policy* 26: 753–771.

Morgan, P. and R. Manning. 1985. Optimal Search. *Econometrica*. 53(4): 923-944.

Mullen, B., C. Johnson, and E. Salas. 1991. Productivity Loss in Brainstorming Groups: A Meta-analytic Integration. Basic Applied Social Psychology. 12(1): 3-23.

Osborn, A. F. 1957. *Applied Imagination*. New York, Charles Scribner's Sons.

Perkins, D. 2000. *The Eureka Effect: The Art and Logic of Breakthrough Thinking*. Norton, New York.

Punj, G. and D. W. Stewart. 1983. Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*. 20: 134-148.

Rao, V. R. and R. Katz. 1971. Alternative Multidimensional Scaling Methods for Large Stimulus Sets. *Journal of Marketing Research*. 13: 488-494.

Ratneshwar, S., L.W. Barsalou, C. Pechmann, and M. Moore. 2001, Goal-Derived Categories: The Role of Personal and Situational Goals in Category Representations. *Journal of Consumer Psychology*. 10(3): 147-157.

Rivkin, J.W. and N. Siggelkow. 2003. Balancing Search and Stability: Interdependencies Among Elements of Organizational Design. *Management Science, 49: 290-311.*

Rivkin, J.W. and N. Siggelkow. 2007. Patterned Interaction in Complex Systems: Implications for Exploration. *Management Science.* 53: 1068-1085.

Rosenkopf, L. and A. Nerkar. 2001. Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry. *Strategic Management Journal*. 22: 287-306.

Rothschild, M. 1974. Searching for the Lowest Price When the Distribution of Prices Is Unknown. *The Journal of Political Economy*, 82(4): 689-711.

Scott, J. 2000. *Social Network Analysis: A Handbook*. Sage Publications.

Seber, G. A. F. 1965. A note on the multiple recapture census. *Biometrika* 52:249-259.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*, 2nd Ed. Griffin.

Simon, H. A. 1996. *The Sciences of the Artificial*, Third Edition, MIT Press, Cambridge, MA.

Singh, J. and L. Fleming. 2009. Lone Inventors as Source of Breakthroughs: Myth or Reality? SSRN working paper: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1299064.

Six, J. M. and I. G. Tollis. 2001. Effective Graph Visualization Via Node Grouping. *Proceedings of the IEEE Symposium on Information Visualization 2001* (INFOVIS'01). 51-58.

Sommer, S.C. and C.H. Loch. 2004. Selectionism and Learning in Projects with Complexity and Unforseeable Uncertainty. *Management Science*, 50(10):1334-1347.

Stevens, G. A. and J. Burley. 1997. 3000 Ideas = 1 Commercial Success! *Research-Technology Management*. 40(3): 16-27.

Stigler, G. J. 1961. The Economics of Information. *The Journal of Political Economy* 69(3): 213-225

Stuart T. E. and J. M. Podolny. 1996. Local Search and the Evolution of Technological Capabilities. *Strategic Management Journal*. 17: 21-38.

Sudman, S., M. G. Sirken, and C. D. Cowan. 1988. Sampling Rare and Elusive Populations. *Science*. 240: 991-996.

Terwiesch, C. 2008. Product Development as a Problem Solving Process, in *Blackwell Handbook on Technology and Innovation Management* (ed. by S. Shane), Wiley-Blackwell.

Terwiesch, C. and K.T. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business Press.

Toubia, O. 2006. Idea Generation, Creativity, and Incentives. *Marketing Science*. 25(5):411-425.

Tversky, A. 1977. Features of Similarity. *Psychological Review* 84 (4): 327-352.

Ulrich, K. T. and S. D. Eppinger. 2008. *Product Design and Development*, 4^{th} Ed. McGraw-Hill Higher Education.

Weitzman, M.L. 1979. Optimal search for the best alternative. *Econometrica* 47 641-654.

Yan, B. and S. Gregory. 2009. Detecting Communities in Networks by Merging Cliques. 2009 *IEEE International Conference on Intelligent Computing and Intelligent Systems* (ICIS 2009). 832–836.

Appendix A: Instructions for Similarity Coding

On the accompanying three paper sheets, you will find a master list of "new ideas" generated as part of an innovation effort.

In this task you will form groups of similar ideas from this list.

First, read through the entire list to become generally familiar with the ideas.

Then, complete two tasks. The first aims to identify *similar* ideas. The second aims to identify *identical* or *essentially identical* ideas. The detailed instructions for these two tasks are provided below. You will record the results of your work in the spreadsheet you've been given.

Before you begin, record in the cells at the top of the spreadsheet your "Lab ID," the "Session Letter" and "Session #" printed at the top of your list. These cells are highlighted in light blue.

Similar Ideas

Consider the list of ideas. **Identify groups of two or more ideas that are that are *similar* to each other.** You should base this grouping on your own notion of similarity. We understand that people think about similarity in their own way, which is fine.

Record the ID numbers for ideas that are similar in the rows in the spreadsheet you've been given (labeled "Similar Ideas"). So, for example, the first group would correspond to Row 4 and the ideas in that group would be entered along Row 4 in Columns B, C, etc. You may find it helpful to give each group a descriptive label in Column A, but this is optional. Feel free to mark up the paper sheet of ideas if that is helpful, but only the information recorded in the spreadsheet will be used in our analysis.

The ideas on your list are drawn randomly from a larger sample, and so it is possible there could be few or many groups of similar ideas.

It is ok to place an idea in more than one group if you wish.

Identical or Essentially Identical Ideas

Consider again the list of ideas and your groups of similar ideas. On the lower portion of the worksheet, **identify groups of two or more ideas that are *identical or essentially identical***.

Record the ID numbers for ideas that are identical or essentially identical in the rows in the spreadsheet you've been given (in the area labeled "Essentially Identical Ideas"). Again, you may find it helpful to give each group a descriptive label, but this is optional.

If ideas are essentially identical, then they are also similar, and so any ideas that appear together in an essentially identical group will also appear together in one or more of your similar groups.

The ideas on your list are drawn randomly from a larger sample, and so it is possible that there could be no ideas on your list that are identical or essentially identical.

Appendix B: Forming Lists of Ideas for Raters

To rate the similarity of ideas as described in Appendix A, we provided subjects with lists of ideas. Ideally, each subject would rate the similarity of all the ideas in an entire data set. However, each of the five data sets had a few hundred ideas, approximately twelve pages of ideas. We saw that it was too hard for people to reliably recall similar and identical ideas over that many ideas. To make the task manageable, we created lists of approximately 75 ideas, or 3 pages of ideas. We used a process to create a set of lists so that (1) every pair of ideas appeared on at least one list and (2) pairs of ideas appeared together on lists an average of about 4 times.

Our algorithm for creating these lists was as follows. Consider every pair of ideas, in random order. If the pair does not appear on any lists, find the shortest list that contains one idea in the pair. Add the other idea to that list. If neither idea appears on any list, add both ideas to the shortest list.

For the data sets with 232 and 249 ideas, we created 40 lists each. For the data sets with 286 to 311 ideas, we created 50 lists each. In total, we had 230 lists of between 68 and 85 ideas that subjects rated for similarity.

Appendix C: Confidence Intervals

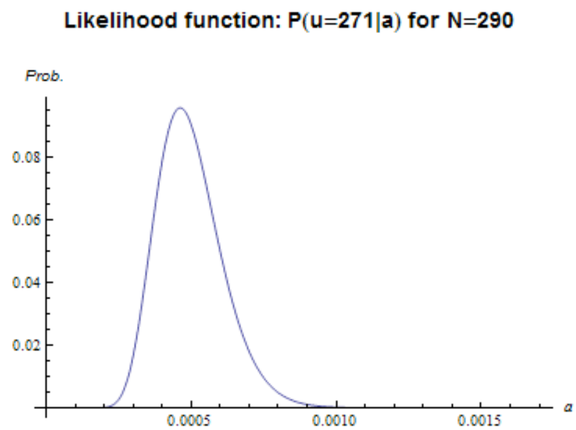
In this appendix, we describe the details of how we computed the confidence intervals on the estimate of the total number of unique ideas, T. We use a Bayesian approach. As such, we derive a posterior distribution $p(a|u)$, i.e., a distribution on the equation parameter (a) given the observed data (the number of unique ideas u in the data set). To do that, we need two components, the likelihood function $p(u|a)$ and the prior distribution $p(a)$.

The likelihood function $p(u|a)$ gives the probability that there are u unique ideas out of N ideas generated, for a particular value of a. The value of $p(u|a)$ is derived from the stochastic process defined by Equation 1: the ith idea is either unique (with probability $p(i) = e^{-ai}$) or not (with probability $1 - p(i)$). The total number of unique ideas out of N ideas generated is therefore the sum of N Bernoulli (i.e., binary 0/1) random variables. Using a central limit theorem (Kallenberg, 1997), we approximate the sum of the Bernoulli random variables as a Normal distribution with mean equal to the expected number of unique ideas $u(N)$ and the variance as the sum of the variances of the Bernoulli random variables. The variance of a Bernoulli random variable with parameter p is $p(1\text{-}p)$. We approximate this sum using the integral

$$\int_0^N e^{-an}\left(1 - e^{-an}\right)dn = \frac{1}{2a}\left(1 + e^{-2aN} - 2e^{-aN}\right)$$

The observation for each data set, the number of unique ideas out of N, is a whole number. The Normal approximation to the sum of the Bernoulli random variables is a continuous approximation. To find the probability that u unique ideas appeared, we use the probability of the Normal random variable being between $u-0.5$ and $u+0.5$.

Below we show an example of a likelihood function for the New Products I data set, with 271 unique ideas out of 290 generated (u(290) = 271). There are a few things to note about the likelihood function. First, it is not a probability distribution; it does not necessarily sum to 1. Second, it is bell-shaped. Values of the parameter (a) around 0.00047 yield 271 unique ideas out of 290 with greater likelihood than values of the parameter that are much lower or much higher. Third, for values of a that are too low or too high, there is essentially no chance that they yield 271 unique ideas out of 300.



For the prior distribution on *a*, we use a "diffuse prior," (Hays and Winkler, 1971, pp. 482-484) representing the case in which the observed information would receive much more weight than the prior. A diffuse prior essentially serves as a uniform distribution on *a* for which we don't have to pre-specify the range. The $p(a)$ is treated as a constant. In our calculations, the range of *a* is effectively narrowed to values of *a* for which $p(u|a)$ is non-zero. (In our numerical analysis, we set the threshold to be 10^{-10}.) (Note: we also checked the case in which the diffuse prior is placed on *T* rather than on *a*. The confidence intervals are shifted up slightly, but are quite similar.)

Putting together the pieces with Bayes' rule, we use $p(u|a)$ to find $p(a|u)$ the probability of *a*, given an observed value of *u*:

$p(a|u) = \frac{p(u|a)p(a)}{\sum_a p(u|a)p(a)}$, which reduces to $p(a|u) = \frac{p(u|a)}{\sum_a p(u|a)}$ because of our assumption that $p(a)$ is constant.

For practical purposes, we discretize the *a* space, looking at values of *a* in intervals of 10^{-5}. For the New Products I data set shown in the figure above, the relevant range for *a* is 0.00012 to 0.00171.

Finally, we use the range of *a* between the 2.5th and 97.5th percentiles of $p(a|u)$ to deduce the corresponding range on *T*.

Appendix D: Test for Statistically Significant Difference of Estimates

To test for the statistical significance of the difference of the estimates for any two data sets, we compute the probability that that difference would be at least as big as observed. The logic is that of a t-test. However, we do not use the t-test per se, because we are testing a difference in the medians of non-Normally distributed quantities, not a difference in means of Normally distributed quantities (as in the t-test).

For each pair of data sets, we simulated 100,000 draws from each median-centered distribution. The distributions are those derived as described in the previous appendix on confidence intervals, the $p(a|u)$. We use the median to center because the point estimate for the model parameter is approximately the median of the distribution. Then we compute the fraction of the simulated pairs that have a difference greater than or equal to the difference in the observed parameter estimates. If very few of the simulated differences are as big as the actual difference, we conclude that it is unlikely that the point estimates (the medians) of the distributions are the same.

Those fractions are shown in Table A1 for each pair of data sets, using the unique counts from the consensus threshold (70% level of agreement).

Table A1: Fraction of 100,000 simulated draws that are greater than or equal to the observed differences between data sets.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
New Ventures					
Web-Based Products	0.57				
New Products I	0.93	0.48			
New Products II	0.89	0.43	0.92		
Classroom Technologies	0.03**	0.01***	0.02**	0.03**	

Appendix E: Multiple Ideas per Person

To examine the question of how much it matters that each person generated five ideas, we run a simulation of the five-unique-ideas-per-person format and see how that format changes the expected number of unique ideas in a data set, compared to the predictions from our baseline model, Equation 2.

For each data set, we simulated q people each generating five ideas. The five ideas are modeled as five draws, without replacement, from a set of T total unique ideas. The q is set determined by Round[N/5]. (Note that there are slight discrepancies with the data: for New Products II, 58 actual participants generated 286 ideas; a few people did not complete all five; therefore we simulated 57 people and use the benchmark $u(285)$.) Table A2 shows the results.

Table A2: Comparison of estimate of N assuming enough ideas are generated by each individual to produce five unique ideas.

	New Ventures	**Web-Based Products**	**New Products I**	**New Products II**	**Classroom Technologies**
N, ideas in data set	232	249	290	286	311
# simulated ideas, 5 Round[N/5]	230	250	290	285	310
Predicted $u(5q)$ from model	218.20	238.91	271.00	266.13	270.25
Average # of unique ideas in 10,000 trials	218.42	239.12	271.31	266.47	270.85

The comparison of the last two rows of this table shows that the restriction that each individual will generate five unique ideas has virtually no effect on the predictions of the model.

Appendix F: Clustering Analysis Accounting for Multiple Ideas per Person

Table A3: A variation of Table 8 in which the random benchmarks reproduce the pattern of multiple ideas per person found in the data. We continue to see support for the hypothesis that opportunities in the data sets are more clustered than random.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Shared-need slice level (1/5)					
Number of clusters, actual	110	133	147	147	99
Number of clusters, random (average over 50 samples)	122	140	159	153	126
T-statistic	4.96***	2.36**	3.66***	1.84*	8.76***
Shared-category slice level (1/10)					
Number of clusters, actual	69	84	88	98	62
Number of clusters, random (average over 50 samples)	81	96	110	105	76
T-statistic	4.83***	4.76***	6.80***	2.17**	6.09***

*** p<0.01, **p<0.05, *p<0.10 two-tailed tests

Appendix G: Sensitivity to Similarity Metric

Table A4: A variation of Table 8, with sensitivity analysis to the scaling factor in the similarity measure. (The original value was 15; here we compare to 17.5, 12.5, and 10.) We continue to see support for the hypothesis that opportunities in the data sets are more clustered than random, especially for the higher values of the scaling factor.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Shared-need slice level (1/5)					
17.5 Actual	106	124	132	131	98
17.5 Random avg	116	132	151	146	123
T-statistic	3.61***	2.94***	6.01***	4.13***	8.67***
15 Actual	110	133	147	147	99
15 Random avg	121	139	159	154	125
T-statistic	3.65***	1.70*	3.61***	2.27**	9.51***
12.5 Actual	114	135	157	151	103
12.5 Random avg	125	145	165	157	127
T-statistic	3.60***	3.28***	2.60**	1.92*	8.20***
10 Actual	122	150	170	159	103
10 Random avg	131	156	176	164	128
T-statistic	2.37**	1.60	1.64	1.31	9.18***
Shared-category slice level (1/10)					
17.5 Actual	64	73	76	88	56
17.5 Random avg	77	89	102	99	74
T-statistic	6.37***	5.49***	10.16***	4.23***	7.07***
15 Actual	69	84	88	98	62
15 Random avg	81	95	110	106	77
T-statistic	5.72***	3.50***	7.68***	2.80***	6.46***
12.5 Actual	74	96	97	105	63
12.5 Random avg	86	102	117	109	78
T-statistic	4.04***	2.01*	6.12***	1.46	6.10***
10 Actual	89	104	111	112	70
10 Random avg	91	112	127	117	80
T-statistic	0.69	2.31**	5.22***	1.96*	3.99***

*** p<0.01, **p<0.05, *p<0.10 two-tailed tests

Table A5: A variation of Table 8, with sensitivity analysis to the functional form of the similarity measure. In this analysis, we do not divide by list length: all pairs on any list are given the same similarity value, except that the longest 5% of lists are excluded, as in the identical analysis. We examined similarity values of 7, 5, and 3. We continue to see support for the hypothesis that opportunities in the data sets are more clustered than random.

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
Shared-need slice level (1/5)					
7 Actual	68	68	76	89	63
7 Random avg	85	93	107	111	87
T-statistic	7.56***	10.24***	13.58***	7.75***	11.00***
5 Actual	83	98	99	107	78
5 Random avg	102	114	129	130	104
T-statistic	6.77***	5.75***	12.21***	9.05***	8.64***
3 Actual	118	143	155	150	96
3 Random avg	125	147	164	157	119
T-statistic	2.76***	1.16	2.46**	1.77*	7.12***
Shared-category slice level (1/10)					
7 Actual	33	36	39	45	27
7 Random avg	49	54	62	65	40
T-statistic	7.48***	8.06***	11.30***	10.22***	7.96***
5 Actual	42	49	51	62	36
5 Random avg	62	70	79	82	54
T-statistic	9.90***	11.21***	14.20***	7.49***	7.90***
3 Actual	70	81	88	98	53
3 Random avg	82	98	110	107	69
T-statistic	5.24***	6.59***	8.17***	2.92***	7.29***

*** p<0.01, **p<0.05, *p<0.10 two-tailed tests