

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

Laura J. Kornish

Leeds School of Business, University of Colorado, Boulder, Colorado 80309, kornish@colorado.edu

Karl T. Ulrich

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, ulrich@wharton.upenn.edu

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.

*Key words:* search; opportunity; opportunities; idea; ideation; idea generation; innovation; creativity; innovation process; opportunity identification; concept development; product development; product design; entrepreneurship

*History:* Received November 9, 2009; accepted July 30, 2010, by Lee Fleming, entrepreneurship and innovation. Published online in *Articles in Advance* November 19, 2010.

## 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, p. 1331).

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, because 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 goals are to answer 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 1,000 more opportunities to be discovered.<sup>1</sup> 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.

This 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.

<sup>1</sup> 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 §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.

### 2.1. 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 2010, 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.

### 2.2. 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 (2008). 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 multidimensional, 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 (1997), 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 “inter-connectedness” 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.

### 2.3. 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, 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.<sup>2</sup> 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:

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

<sup>2</sup> 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.

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.)

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

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 §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 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.

#### 4.1. Ideas

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.

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 midcareer 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

**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)?

issues in our analysis (§§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 to which it has access. 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 yield 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.

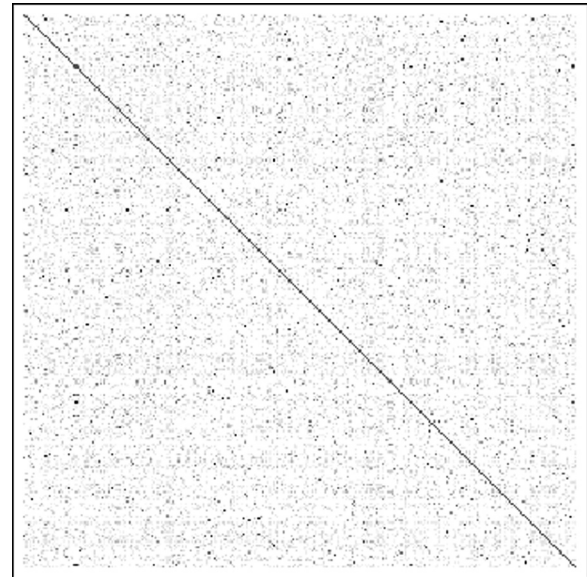
#### 4.2. Quality Measures

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.

#### 4.3. 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.

Figure 1 Similarity Between Pairs of Ideas for the New Products I Data Set



Notes. 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 ideas  $i$  and  $j$ . In this data set, approximately 26% of the pairs have nonzero similarity.

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  $(290 \times 289)/2 = 41,905$  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 ideas  $i$  and  $j$ . 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 pretests, 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.

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, and 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 multidimensional 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.

## 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.

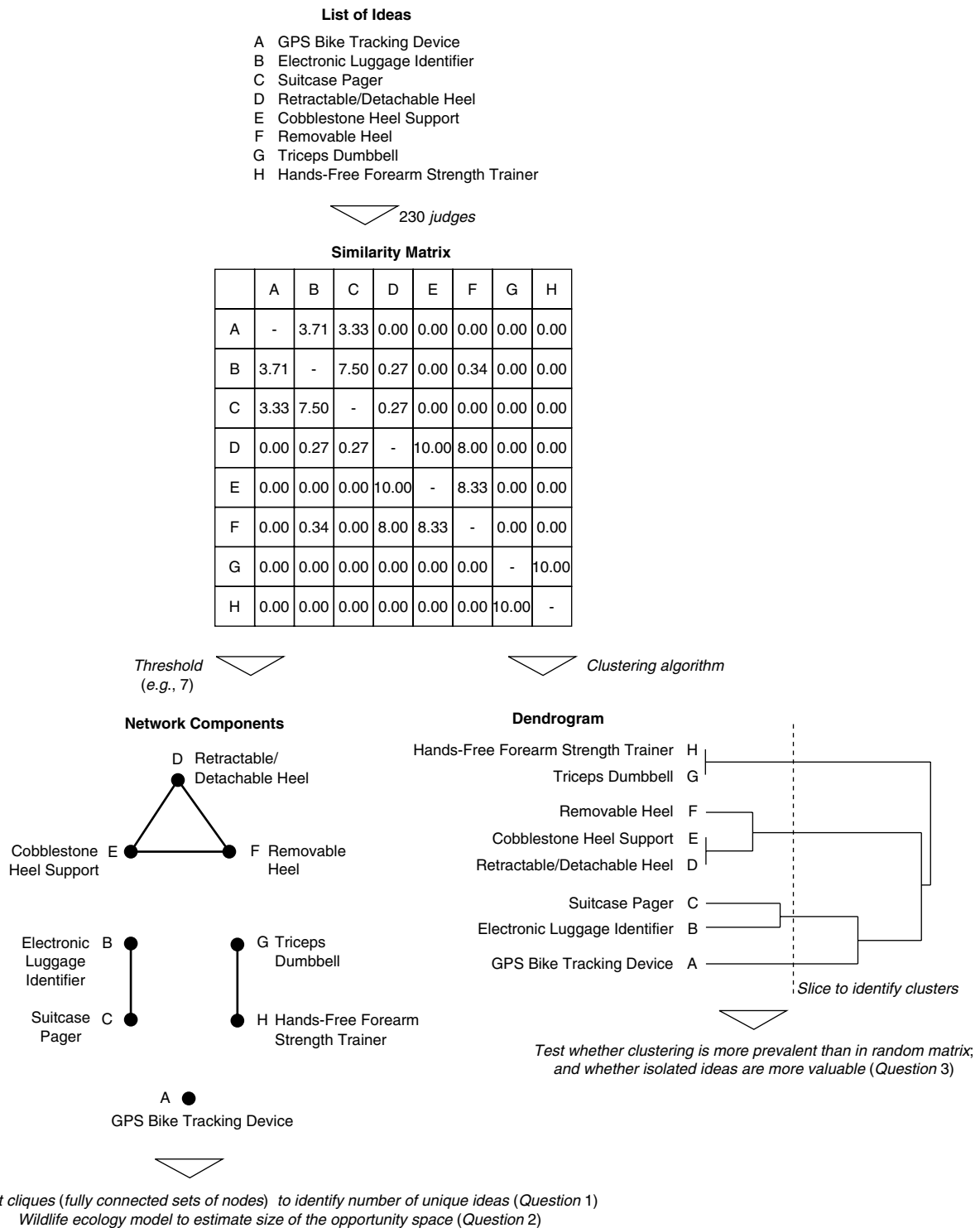
### 5.1. Determining the Number of Unique Ideas

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 cutoffs 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

Figure 2 Analytical Framework and Approach



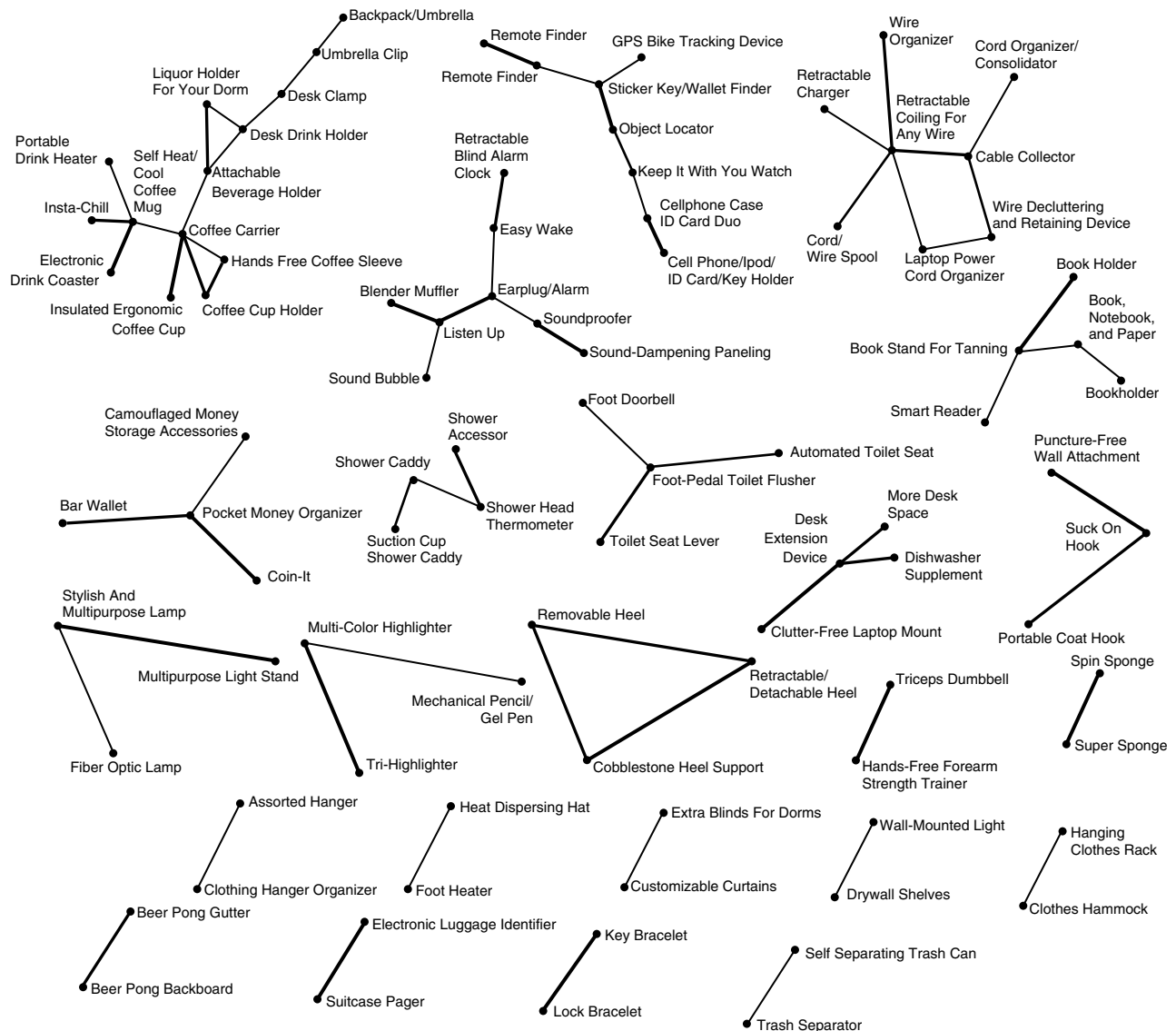
Note. This analysis is performed for each of five independent data sets.

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  $49 \times 48/2 = 1,176$  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

**Figure 3** Clusters of Identical Ideas for the New Products I Data Set Based on the Majority Threshold for the Definition of Identical



*Notes.* 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.

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 24 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



**Table 2** Distribution of Network Component Sizes for Each Data Set and for Two Thresholds Defining *Identical*

	New Ventures		Web-Based Products		New Products I		New Products II		Classroom Technologies	
	<i>N</i>	Fraction of ideas (%)	<i>N</i>	Fraction of ideas (%)	<i>N</i>	Fraction of ideas (%)	<i>N</i>	Fraction of ideas (%)	<i>N</i>	Fraction of ideas (%)
Panel A: Majority threshold for identical ( $\geq 50\%$ of raters identify pair as identical)										
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 ( $\geq 70\%$ of raters identify pair as identical)										
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

Note. 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).

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 multidimensionality 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.

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 ( <i>N</i> )	232	249	290	286	311
Number of unique ideas ( <i>u</i> ) at majority threshold	191	216	252	231	216
Percent unique (%)	82	87	87	81	69
Number of unique ideas ( <i>u</i> ) at consensus threshold	220	238	271	267	271
Percent unique (%)	95	96	93	93	87

## 5.2. 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<sup>3</sup> to  $u = (1/a)(1 - e^{-aN})$ . 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 gener-

<sup>3</sup>Dawkins (1991) gives an approximation to *T* as  $u^2/(2(n - u))$ . For the first four data sets, this approximation underestimates *T* by about 10%; for the fifth one, it underestimates by nearly 20%.

**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	2,165	2,735	2,115	2,056	1,103
Lower bound for $T$ (2.5th percentile)	1,205	1,493	1,333	1,299	806
Upper bound for $T$ (97.5th percentile)	3,704	4,762	3,333	3,226	1,493

ated increases. Different domains and generating processes would exhibit different curves.

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.

**5.3. Confidence Intervals**

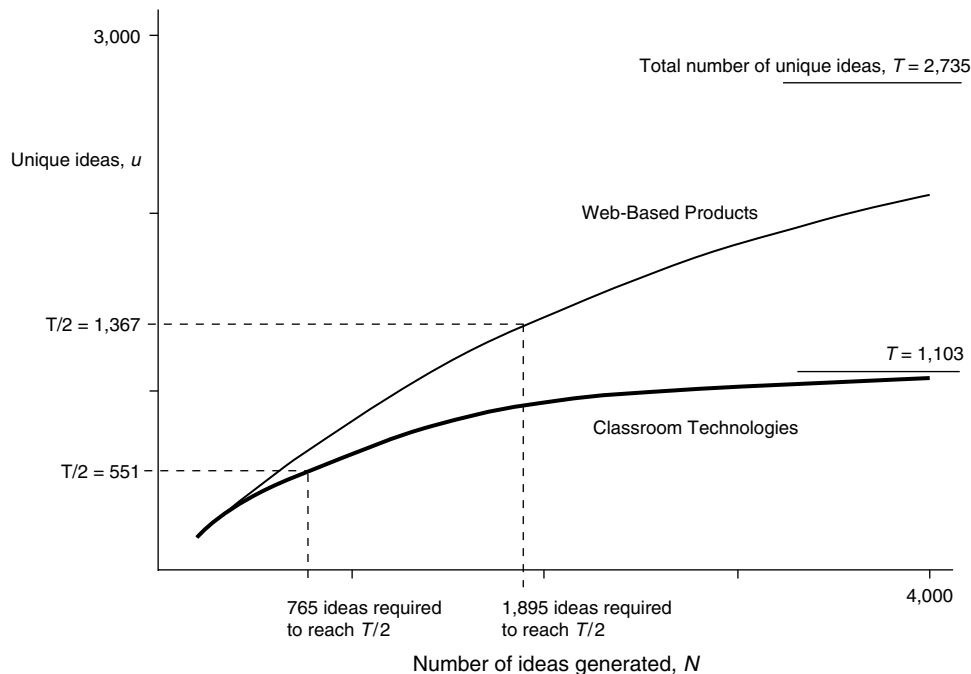
Using our model, we have derived point estimates of the total number of unique ideas,  $T$ , for each data set. Our model for the probability that the next idea is unique (Equation (1)), dictates a stochastic process for the number of unique ideas in any set. Using that uncertainty, we can numerically 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

**Figure 4** Number of Unique Ideas,  $u$ , Expected for a Given Number of Ideas Generated,  $N$



*Note.* Two domains are shown, Web-Based Products and Classroom Technologies.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

**Table 5** Estimates of the Total Number of Unique Ideas,  $T$ , Based on Empirical Relative Frequency of Ideas

	New Ventures	Web-Based Products	New Products I	New Products II	Classroom Technologies
$T$ assuming each idea equally likely	2,165	2,735	2,115	2,056	1,103
$T$ assuming empirical frequency distribution	2,268	2,839	2,334	2,205	1,192

*Note.* These estimates use the consensus threshold for identical, 1,000 simulation trials, and a grid search interval of 15.

every single idea. In contrast, the innovation charges for the other domains were more abstract, soliciting ideas for general products and ventures.

#### 5.4. 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 to 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 nonuniform distribution raises the estimate somewhat.

#### 5.5. 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.

### 6.1. 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 5% 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.

**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

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.<sup>4</sup> 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.<sup>5</sup> When averaged across all raters, the weighted similarity score exhibits a relatively smooth unimodal distribution, skewed toward 0, and with a thin tail stretching to the maximum value of 10.

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

## 6.2. 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, that 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. Also, 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.

## 6.3. 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

<sup>4</sup> 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.

<sup>5</sup> Table 8 refers to more details of this sensitivity analysis.





















