MANAGEMENT SCIENCE
Vol. 56, No. 4, April 2010, pp. 591–605
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5604 | 0591

informs®
DOI 10.1287/mnsc.1090.1144

Idea Generation and the Quality of the Best Idea

Karan Girotra
Technology and Operations Management, INSEAD, 77305 Fontainebleau, France,
karan.girotra@insead.edu

Christian Terwiesch, Karl T. Ulrich
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
{terwiesch@wharton.upenn.edu, ulrich@wharton.upenn.edu}

In a wide variety of settings, organizations generate a number of possible solutions to a problem—*ideas*—and then select a few for further development. We examine the effectiveness of two group structures for such tasks—the *team* structure, in which the group works together in time and space, and the *hybrid* structure, in which individuals first work independently and then work together. We define the performance of a group as the quality of the *best* ideas identified. Prior research has defined performance as the average quality of ideas or the number of ideas generated, ignoring what most organizations seek, a few great ideas. We build a theory that relates organizational phenomena to four different variables that govern the quality of the best ideas identified: (1) the average quality of ideas generated, (2) the number of ideas generated, (3) the variance in the quality of ideas generated, and (4) the ability of the group to discern the quality of the ideas. We test this theory with an experiment. We find that groups organized in the hybrid structure are able to generate more ideas, to generate better ideas, and to better discern the quality of the ideas they generate. Moreover, we find that the frequently recommended brainstorming technique of building on others' ideas is counterproductive; teams exhibiting such buildup neither create more ideas, nor are the ideas that build on previous ideas better.

Key words: creativity; brainstorming; innovation; idea generation; idea selection; team; group; individuals; nominal group; interactive buildup
History: Received December 16, 2007; accepted December 9, 2009, by Christoph Loch, R&D and product development. Published online in *Articles in Advance* February 24, 2010.

1. Introduction

Virtually all innovation processes include generating and selecting opportunities or *ideas*. When a movie studio creates a new feature film, it typically considers several hundred plot summaries, a few of which are selected for further development. When a company develops the branding and identity for a new product, it creates dozens or hundreds of alternatives and picks the best of these for testing and refinement. When a consumer goods firm develops a new product, it typically considers many alternative concepts before selecting the few it will develop further. Generating the ideas that feed subsequent development processes thus plays a critical role in innovation.

The success of idea generation in innovation usually depends on the quality of the *best* opportunity identified. For most innovation challenges, an organization would prefer 99 bad ideas and 1 outstanding idea to 100 merely good ideas. In the world of innovation, the extremes are what matter, not the average or the norm (Dahan and Mendelson 2001, Terwiesch and Loch 2004, Terwiesch and Ulrich 2009). This objective is very different from that in, for example, manufacturing, where most firms would prefer to produce 100 units with good quality over making 1 unit with exceptional quality followed by 99 that have to be scrapped.

When generating ideas, an organization makes choices by intention or default about its creative problem-solving process. In this paper, we investigate two commonly suggested organizational structures for groups generating ideas. In the first, individuals work together as a *team* in the same time and space. The team approach is widely used in organizations (Sutton and Hargadon 1996). Despite its wide usage, hundreds of experimental studies have criticized team processes as relatively ineffective (cf. Diehl and Stroebe 1987, 1991). In the second approach, individuals work independently for some fraction of the allotted time (sometimes called a *nominal group*) and then work together. This *hybrid* process has been suggested and studied in the prior literature as a way of effectively combining the merits of individual and team approaches (cf. Robbins and Judge 2006, Paulus et al. 1996, Stroebe and Diehl 1994). These studies find that the hybrid approach leads to more ideas and to higher satisfaction with the process among participants. Although the two approaches have been compared for many years (Pugh 1981), the prescriptions in the literature conflict.

The existing idea generation literature (often called the *brainstorming* literature) exhibits three gaps with respect to idea generation in innovation management. First, most papers focus on the number of ideas generated, as opposed to their quality, with the tacit assumption that *more* ideas will lead to *better* ideas. Second, the few papers that look at the quality of ideas look at the *average* quality of ideas as opposed to the quality of the *best* ideas. Third, the focus of the existing literature is entirely on the *creation* process and ignores the *selection* processes that groups use to pick the most promising ideas for further exploration.

Given our focus on the use of idea generation in innovation, our metric for effectiveness is the quality of the ideas selected as the *best*. Building on prior work on innovation tournaments and on extreme value theory applied to innovation, we build a theory that relates organizational phenomena to four different variables that govern the underlying statistical process of idea generation and selection: (1) the *average quality of ideas generated*, (2) the *number of ideas generated*, (3) the *variance in the quality of ideas generated*, and (4) the *ability of the group to discern the quality of the ideas*. Each of these variables affects the quality of the best ideas produced by a team or a hybrid group.

We report on a laboratory experiment, which compares the two group structures with respect to each of these four variables individually, and which measures their collective impact on the quality of the best idea. An accurate measurement of idea quality is central to our work. Whereas most prior research has relied on the subjective evaluation of idea quality by one or two research assistants, we use two alternative approaches: a Web-based quality evaluation tool, which collects about 20 ratings per idea, and a purchase-intent survey, which captures about 40 consumer opinions about their intent to purchase a product based on the idea. Our framework, with its emphasis on the importance of the best idea, and our novel experimental setup let us make the following three contributions.

1. We find evidence that the best idea generated in the hybrid structure is better than the best idea generated by a team. This result is driven by the fact that the hybrid structure results in about three times as many ideas per unit of time, and that these ideas have significantly higher average quality.

2. We find that the hybrid structure is better at identifying the best ideas from the set of ideas it previously generated. However, we also find that both team and hybrid structures are, in absolute terms, weak in their ability to assess the quality of ideas.

3. We show that idea generation in teams is more likely to lead to ideas that build on each other. However, in contrast to the common wisdom articulated by many proponents of team brainstorming, we show that such buildup does not lead to better idea quality. In fact, we find that ideas that build on a previous idea are worse, not better, on average.

The remainder of this paper is organized as follows. We review the relevant literature in §2. We then develop our theoretical framework in §3 and state our main hypothesis. Section 4 describes the experiment, and §5 describes our performance measures. Section 6 reports our main results. Section 7 looks at the role of buildup in groups, and §8 contains concluding remarks.

2. Literature

The role of organizational processes in idea generation has been examined in the social psychology literature and in the innovation management literature. The social psychology literature has examined the idea generation process in detail in what is often called the *brainstorming* literature. The innovation management literature has focused on innovation outcomes and organizational forms.

The social psychology literature mostly originated with Osborne's (1957) book, *Applied Imagination*, which introduced the term *brainstorming*. Osborne (1957) argued that working in teams leads to multiple creative stimuli and to interaction among participants, resulting in a highly effective process. His argument spawned many experiments. Diehl and Stroebe (1987) and Mullen et al. (1991) provide a detailed overview of this literature. These studies experimentally examined groups generating ideas as teams or as individuals. In terms of performance metrics, the literature focuses on the *average quality of the ideas generated*, the *number of ideas generated*, and measures that combined the two, such as the *total quality produced*. Quality ratings for ideas generated are typically provided through evaluations by research assistants. For example, in Diehl and Stroebe (1987), the ideas were rated by one research assistant, and a second assistant was used to verify the reliability. 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.) Given the focus of these studies on productivity and average quality, they conclude that team processes are inferior to individual processes. However, this main conclusion is in stark contrast with Osborne's (1957) hypothesis, empirical work that examines metrics such as variance of the quality distribution (Singh and Fleming 2010), and anecdotal evidence that team idea generation processes (i.e., brainstorming) are widely used in organizations.

In line with the social psychology literature, we also conduct experiments. However, in contrast to this literature, we examine idea generation in the specific context of generating ideas in response to an innovation challenge. Given the focus on innovation, we are concerned with the quality of the *best* ideas resulting from the idea generation process, not with the average quality. Furthermore, we depart from this literature by employing a novel method of evaluating idea quality based on a large panel of independent raters and on a purchase-intent survey conducted with subjects from the target market segments.

To resolve the contrast between the social psychology literature and anecdotal evidence about the practices of real organizations, Sutton and Hargadon (1996) conducted a field-based observational study of the product design consulting firm IDEO. They found that contextual differences between the lab and the real world (e.g., the nature of problems addressed) may explain the contrast between practice and the laboratory findings. More recently, Kavadias and Sommer (2009) took an innovative approach to this paradox. They showed analytically that the specific nature of the problem and group diversity matter, and they conjectured that the experimental evidence may be an artifact of exploring simple idea generation problems that are not representative of real situations.

The role of organizational structure in the idea generation process has also been examined with large-scale empirical studies, most notably by Singh and Fleming (2010), who used patent data to study differences in quality variance between inventors who work by themselves and those who collaborate. They examined the impact of collaboration on both the upper and lower tails of the quality distribution. Quality is measured as the number of citations received by the patent. Taylor and Greve (2006) also examined average quality and variance of creative output in the comic book industry. They measured quality as the collector-market value of a comic. Singh and Fleming (2010) found an asymmetric impact of collaboration, in that collaboration reduces the number of bad ideas and increases the number of very good ideas. Taylor and Greve (2006) found that quality variance is lower for individual innovators. In the experimental studies mentioned above, the differential resource investment between individuals and teams can be controlled by aggregating individual innovators into synthetic teams (also called *nominal groups*); this is impossible to do in natural empirical studies. Thus, it is hard to draw conclusions about productivity from these studies, though the results on average quality and variance directly inspire our work.

Last, the statistical view of innovation, which is at the core of our analyses and hypotheses, was first developed by Dahan and Mendelson (2001). They modeled creation as a series of random draws from a distribution followed by a selection from the generated ideas. We employ this model to identify the statistical properties that influence the quality of the best idea. We summarize the relevant literature and the key differences between the literature and our study in Table 1.

3. Theory

Our theoretical framework has two parts. First, we provide a statistical model of the idea generation and selection process, which is comprised of four explanatory variables. Second, we use the prior literature on group decision making to link group structure to each of these four variables. In this way, we connect group structure to idea generation performance.

In the context of innovation, the goal of idea generation rarely is to maximize the number of ideas, nor is it to maximize the average quality of ideas. Instead the goal is usually to maximize the quality of the *best idea* or the few best ideas. As we argued in the Introduction, most organizations would prefer to generate 1 incredibly good idea and 99 terrible ideas rather than generating 100 merely decent ideas.

Prior research on innovation has modeled the process of idea generation as repeated sampling from an underlying quality distribution (Dahan and Mendelson 2001, Terwiesch and Loch 2004). This literature draws on a branch of statistics known as *extreme value theory*, which shows that the maximum value of a sample of n draws from an underlying distribution F can be modeled as a function of three variables: the *sample size*, the *mean of the underlying distribution*, and the *variance of the underlying distribution*.

In the innovation context, if F reflects the distribution of quality of the ideas available to the group, and if the idea generation process is essentially repeated independent draws from this distribution, then the expected quality of the *best* ideas is driven by the *number of ideas generated by the group*, the *average quality of the underlying quality distribution*, and the *variance of the underlying quality distribution*. These three variables are illustrated in Figure 1. These insights and arguments are presented as formal proofs in the appendix, §A.1.

This extreme value logic assumes that a group can discern good ideas from bad ideas. However, a group might generate exceptionally good ideas, but incorrectly characterize those ideas. Thus, the ability of a group to discern idea quality is also critically important in determining performance. This fourth factor is shown on the right side of Figure 1, using the illustrative case of a group selecting the *third-best* idea generated rather than the very best idea.

Of course, in the innovation context, the quality of a raw idea is clouded by a great deal of uncertainty, and

Table 1 Summary of Literature with Comparison to This Study

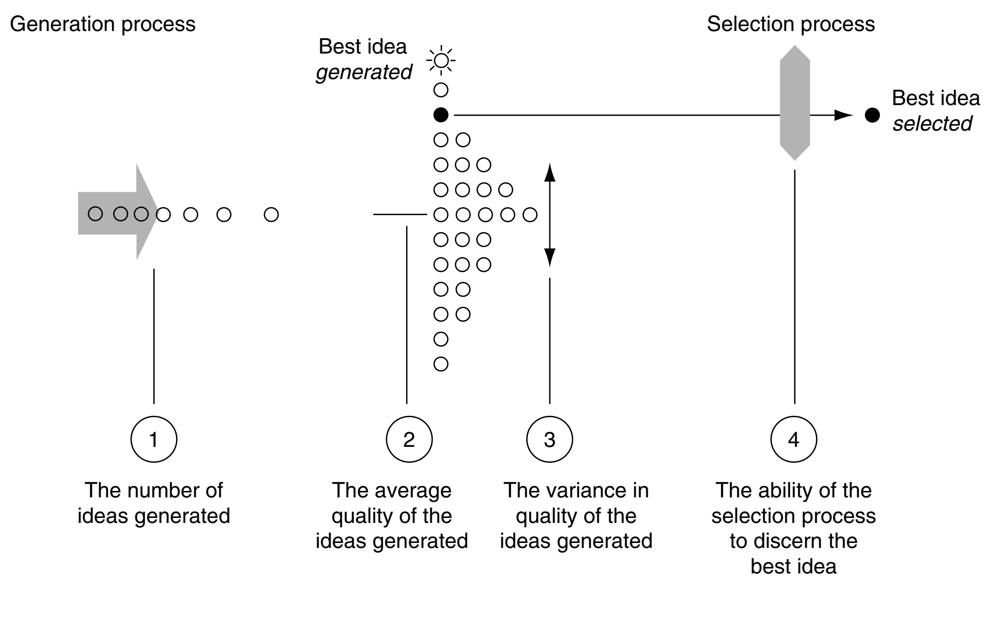
Research	Setting/Methodology	Measure of idea quality	Metrics	Results
Osborne (1957)				Introduced brainstorming
Social psychology literature, summarized by Diehl and Stroebe (1987, 1991), Stroebe and Diehl (1994)	Lab, experimental	Rating by an assistant (second assistant used for reliability), rating by an expert	Mean quality and productivity	Productivity: Individual > teams Mean quality: Equivocal results No reason to work in teams!
Sutton and Hargadon (1996)	Industry (IDEO), observational			Contextual differences between lab and the real world
Taylor and Greve (2006)	Comic book industry, empirical	Collector market value of a comic	Mean quality and variance	Variance: Teams > Individuals Moderating effects of knowledge diversity, team experience, workloads, tenure, organizational resources
Singh and Fleming (2010)	Patent data, empirical	No of patents, citations (use of patent)	Mean quality, variance, and productivity	Mean: Teams > Individuals Variance: Individuals > Team
Kavadias and Sommer (2009)	Analytical			Depends on problem structure and team diversity (experience and knowledge)
Dahan and Mendelson (2001)	Analytical	Best idea (extreme value)	Extreme value of quality	
Present study	Lab (with trained subjects), experimental	Business value and purchase intent based on large panel of judges	Mean quality, variance, productivity, ability to discern best ideas, quality of *best* idea	Reported in §§6 and 7

so we should think of idea quality not as a deterministic quantity, but rather as the expected value of the idea assuming optimal further allocation of resources. The uncertainty associated with idea quality is one of the reasons organizations typically invest in a few of the best ideas rather than in only one idea.

3.1. Linking Group Structure to Process Variables

This paper concerns the organizational structure of groups engaging in idea generation, and so here we link existing organizational theory to the four factors articulated in our model of the idea generation process. Based on these theoretical pathways we can

Figure 1 Four Factors Underlying the Performance of the Idea Generation Process



then hypothesize how group structure should influence idea generation performance.

Recall that we focus on two basic structures for group idea generation. In the *team* structure, a group works together interactively in the same time and space, essentially sharing a common experience based on the same information. In the *hybrid* structure, a group first works independently with no interaction of any kind and with each individual potentially accessing different knowledge and navigating the problem-solving challenge in different ways. The hybrid structure includes a second phase in which the individuals work together to share their findings from the individual phase and to perform additional exploration together. How might individual work in combination with working together offer advantages over a pure team structure?

Number of Ideas. A large body of research has shown that individuals generate more ideas than groups, normalizing for person-time devoted to the task. Diehl and Stroebe (1987) provide a review and summary of this literature. The literature cites three main underlying causes of this phenomenon. First, *free riding* is the standard agency problem encountered when output is measured collectively as opposed to individually. Second, *evaluation apprehension* refers to an inhibition to express ideas because of the fear of a negative reaction from other group members. Third, *production blocking* occurs when one person in a group is speaking while the others are waiting to speak, effectively limiting the number of ideas that can be produced to the number that can be articulated orally by one group member speaking at one time. These mechanisms unambiguously suggest that hybrid structures, because of their individual phase of work, should perform better than team structures in terms of the raw *number of ideas generated* per unit of person time invested.

Average Quality of Ideas. Just as free riding can be expected to decrease team productivity, it could also decrease the quality of ideas generated by teams (Wageman 1995), and indeed this effect has been shown experimentally (Diehl and Stroebe 1987). However, advocates of team-based idea generation (e.g., brainstorming) have long argued that teams generate better ideas than individuals working separately because interactions between people in a team are synergistic. One person can build on the ideas of another in a way that increases the *quality of the ideas*. As far as we know there has not been any theoretical or empirical support for the claim that these ideas are better than ideas that are generated independently. This is an issue that our data will allow us to address.

Variance in Quality of Ideas. Ideas that are similar in content and approach are likely to also be similar in quality. This is because idea quality in innovation should be related to some combination of the need the idea addresses and the solution the idea embodies. If two ideas address a similar need or embody a similar solution, they are likely to be of similar quality (Kornish and Ulrich 2009). Two phenomena are likely to limit the variation in the ideas generated by teams. First, if individuals working in teams indeed build on the ideas of others, then those ideas are likely to be similar to one another (Kavadias and Sommer 2009). Indeed, we expect people in the individual phase of the hybrid process to explore more broadly than those that work together in a team (Singh and Fleming 2010). Second, teams are likely to exhibit conformity effects. Janis (1982) finds that teams place strong pressures on their members to conform to one another, although those pressures may be largely subconscious. Teams also minimize internal conflict and focus on issues that maximize consensus (Van de Ven 1986). This pressure to conform is likely to suppress variation in the substance of ideas, and therefore in the quality of ideas. Thus, we expect hybrid structures to exhibit more *variance in idea quality* than team structures.

Ability to Discern the Best Idea. Combining the views of several different group members to assess the value of an idea should result in better estimates than those made by a single individual. However, the structure of the group may influence this accuracy. Although we know of no prior literature that specifically addresses the issue of how the structure of groups might influence the ability to discern quality, we theorize that two effects may be at work. First, the individual phase of the hybrid structure forces individuals to be highly engaged in the problem-solving task. This arousal may increase the acuity of their judgments when they evaluate the quality of ideas in the group phase of their work. Second, judgments of ideas may be path dependent in the following way. Most of the ideas generated in the hybrid process are likely to be generated in the individual phase. As a result, most of the ideas reviewed in the group phase of work will appear out of nowhere without information on the path by which the idea was generated. In contrast, for the team structure, all group members experienced together the path by which each idea was generated, and so they may respond not simply to the idea but to its history. We conjecture that evaluating an idea without knowledge of the specific trajectory of its origin may result in more accurate judgments than when the evaluator has experienced the

Figure 2 Organizational Arguments for How a Hybrid Structure Influences the Four Factors Underlying Performance

	1	2	3	4
Statistical drivers of the quality of the best Idea	The number of ideas generated	The average quality of the ideas generated	The variance in quality of the ideas generated	The ability of the selection process to discern the best idea
Arguments against team structure in favor of hybrid structure	Free riding Evaluation apprehension Production blocking	Free riding Interaction and build up of ideas (typically argument for teams)	Build up of ideas Group conformity	Limited engagement of team members Path dependence

complete process by which the idea was generated. Therefore, we expect the hybrid structure to result in more accurate judgments of idea quality than the team structure.

Figure 2 summarizes these organizational factors and their relation to the four basic variables in our statistical model of performance. Based on these factors, we expect groups adopting the hybrid process to perform better than teams in terms of the quality of the best ideas they produce, which leads to this central hypothesis:

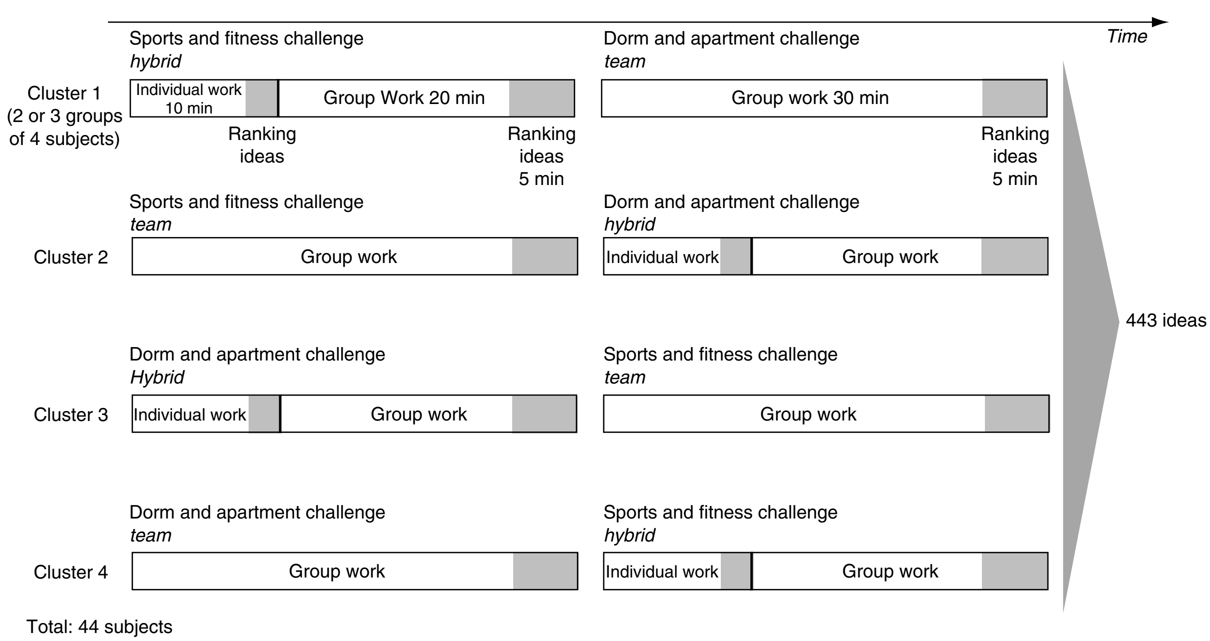
Hypothesis 1. *The quality of the best ideas generated and selected by a hybrid group is higher than that of a team.*

In addition to testing this overarching hypothesis, we also aim to measure empirically the impact of group structure on each of the four variables in the statistical model of the idea generation process, and thereby unpack the causes of differences in performance between the two structures.

4. Experimental Design

To compare the performance of hybrid groups and teams, we designed an experiment that allowed us to manipulate the organizational structure and measure the *number of ideas generated*, the *mean quality of the ideas*, the *variance in quality of the ideas*, and the *ability of groups to discern idea quality*. We employed a within-subjects design for this study. In such a design, each subject generates ideas under *both* the treatments—team and hybrid. Such a design allows us to control for the substantial differences in individual ability, group composition, and group dynamics. Furthermore, with this design the within-group variance in idea quality can be separated from the across-team variance. Figure 3 illustrates the experimental design,

Figure 3 Experimental Design



showing the tasks over time for each of four separate clusters of subjects.

4.1. Subjects

Subjects for the experiment were recruited from an upper-level product design elective course at the University of Pennsylvania. All subjects had participated in multiple brainstorming and idea generation exercises prior to the experiment and had received training in idea generation techniques. The 44 subjects came from a wide variety of majors, with a majority in engineering and business. Most subjects were juniors, seniors, or master's degree candidates. All experiments were conducted after obtaining prior approval from the human subjects committee at the university, and participation in the exercise was voluntary and had no bearing on performance in the course. The subjects were informed that this was as an experiment to understand the idea generation process. Because extrinsic incentives are known to limit creative behavior (Amabile 1996), no explicit incentives or compensation were provided for participation or performance in the experiment. The subjects appeared highly engaged in the experiment and seemed to enjoy addressing the challenges.

4.2. Treatments

In the team treatment, subjects were divided randomly into groups of four. Each team was given 30 minutes to complete an idea generation challenge. The subjects were asked to record each idea on a separate sheet of paper. A prestapled and preordered bundle of sheets was provided to each team. The sheets included an area for notes related to the idea and a designated area to record a title and a 50-word description. At the end of 30 minutes, the subjects were given an additional five minutes and instructed to develop a consensus-based selection and ranking of the five best ideas generated by their team.

In the hybrid treatment, subjects were asked to work individually on an idea generation challenge for 10 minutes. At the end of 10 minutes, the individuals were asked to rank their own ideas. The subjects were then divided randomly into groups of four and given an additional 20 minutes to share and discuss their ideas from the first phase and to develop new ideas. All ideas from both the individual and group portion of the process were recorded on sheets as described for the team process. At the end of the group phase of the hybrid treatment, subjects were given an additional five minutes and instructed to develop a consensus-based selection and ranking of the five best ideas generated by their group.

4.3. Experiment

Participants were divided into four clusters of either two or three groups of four subjects each. Two clusters were administered the hybrid treatment first followed by the team treatment, and the other two were administered the team treatment first followed by the hybrid treatment. Half of the clusters were given Challenge 1 first; the other half were given Challenge 2 first. The paths for the four clusters are illustrated in Figure 3. This setup allowed us to control for effects arising out of the order of treatments, the order of the challenges, and/or those related to interactions between the treatments and the challenges. The two challenges were as follows.

Challenge 1. You have been retained by a manufacturer of sports and fitness products to identify new product concepts for the student market. The manufacturer is interested in any product that might be sold to students in a sporting goods retailer (e.g., City Sports, Bike Line, EMS). The manufacturer is particularly interested in products likely to be appealing to students. These products might be solutions to unmet needs or improved solutions to existing needs.

Challenge 2. You have been retained by a manufacturer of dorm and apartment products to identify new product concepts for the student market. The manufacturer is interested in any product that might be sold to students in a home-products retailer (e.g., IKEA, Bed Bath and Beyond, Pottery Barn). The manufacturer is particularly interested in products likely to be appealing to students. These products might be solutions to unmet needs or improved solutions to existing needs.

A total of 443 ideas were generated. A sample of ideas generated is provided in the appendix, §A.2.

5. Measurement of Performance

Because an accurate measurement of idea quality is essential to testing our theory, we estimated the quality of the ideas two different ways: the business value of the product idea and purchase intent. We believe that these methods are much more accurate than those used in prior studies. We also measured the extent to which an idea builds on the idea that preceded it.

5.1. Business Value

First, we estimated the utility of the ideas to a commercial organization that might develop and sell the products. To do this, we assembled a panel of 41 MBA students, completely distinct from subjects involved with the first phase of the experiment, who had all received formal training in the valuation of new products through a series of graduate classes. This panel was asked to assess the business value of the generated product ideas using a scale from 1 (lowest value) to 10 (highest value). The ideas were presented independently and sequentially to the panelists in a random order using a Web-based interface. Each panelist rated between 206 and 237 different ideas. Each idea was rated by at least 20 different members of the panel. To verify the reliability of these ratings, we follow the method prescribed by Gwet (2002). We constructed Kappa (8.99, 2.92) and AC1 (13.38, 7.59) statistics for each of the two idea domains.[1] All statistics suggest very high levels overall reliability in classification of ideas on our 10-point scale.

5.2. Purchase Intent

We also evaluated the product ideas from the perspective of potential consumers. For this exercise we enrolled 88 subjects from the target market for the product ideas generated—college students. The participants in the survey were provided descriptions of the product ideas and were asked to assess on a 10-point scale their likelihood of purchasing the products. The product descriptions were provided in a randomized order, and each survey participant saw between 200 and 245 different ideas. Each idea was rated by at least 44 different potential customers using a Web-based interface and following standard market research techniques for measuring purchase intent (cf. Ulrich and Eppinger 2007, Jamieson and Bass 1989). To verify the reliability of the ratings, we again follow the method prescribed by Gwet (2002). We constructed Kappa (11.45, 9.93) and AC1 (8.92, 11.627) statistics for each of the two idea domains. All statistics suggest very high levels of overall levels of reliability in classification of ideas on our 10-point scale.

Previous research has characterized the quality of new products as multidimensional, including the dimensions of attractiveness and feasibility. We also created a multidimensional quality scheme composed of five different metrics: *technical feasibility* (to what extent is the proposed product feasible to develop at a reasonable price with existing technology), *novelty* (originality of the idea with respect to the unmet need and proposed solution), *specificity* (the extent to which the idea included a proposed solution), *demand* (reflecting market size and attractiveness), and *overall value*. To rate ideas on these dimensions, we recruited a team of two graduate students specializing in new product development and instructed them to rate each idea with respect to these dimensions on 10-point scale. We discarded all ratings for which the two raters disagreed by more than two points. Looking at the remaining ratings, we found that the five dimensions were highly correlated. Factor analysis suggested using only one composite factor for the five metrics. Furthermore, each of the metrics was highly correlated with estimates of business value and probability of purchase which we constructed using larger panels. In light of this correlation and the apparent lack of independent underlying dimensions in the expert judgments, we will present our results using the business value and purchase intent from the two large panels of judges.

5.3. Buildup of Ideas

The phenomenon of progressive buildup of ideas is frequently brought up by advocates of team processes like brainstorming. In our study, we directly explore the role of buildup. To do this, we hired three independent judges to code the substance of ideas on different dimensions. Ideas generated in Challenge 1, sports and fitness, were categorized along the following three dimensions: the type of product, the principal sporting activity associated with the product, and the key benefit proposition of the proposed product. The coders were provided with a set of exhaustive and mutually exclusive potential categorizations for each of the three dimensions. These categories were developed by examining product classifications by the online retailers Amazon, Wal-Mart, and Buy.com. Unrepresented categories in the data were eliminated. As an example, the product idea "cleated shoe covers—a protection for shoes with cleats, to enable walking on hard surfaces without damaging the cleats," was categorized by our coders as footwear (type of product), field sports (principal sporting activity), and convenience (key benefit proposition). The full list of categories for each of the three dimensions is provided in the appendix, §A.3.1.

Ideas generated in Challenge 2, products for a dorm or apartment, were categorized in a similar manner. The corresponding dimensions were product category, the typical room or location of that product, and the key benefit. The full list of categories for ideas generated for Challenge 2 is in the appendix, §A.3.2.

To construct our buildup metric, we compare the classification of two consecutively generated ideas. For example, if the idea shares all three dimensions with the preceding idea, it earns a buildup score of 3. More generally, the buildup score is the number of dimensions on which an idea shares a value with the idea generated immediately previously. We average this buildup score across the three independent judges.

6. Results

We first report results on the comparisons of the four key variables in our statistical framework: (a) the

[1] The Kappa statistic was first developed by Cohen (1960) as a statistical measure of the interrater agreement for categorical data. More recent developments such as that by Gwet (2002) question this statistic and suggest using the first-order agreement coefficient, or the AC1 statistic. For both statistics, a value statistically greater than 0, indicates interrater agreement. Estimates of these statistics are distributed normally, and above we report the z-score associated with estimates of these two statistics. The high z-scores suggest that the observed interrater agreement is very unlikely to arise out of random chance, and that indeed the rating panel rated ideas in an internally consistent fashion.

Table 2 Results Comparing Team and Hybrid Treatments for Each of the Dependent Variables

Discussion section	Statistic compared	N	F-statistic for team/hybrid[c]	Least square mean estimate for hybrid[d]	Least square mean estimate for team[d]	Difference of least square means: Hybrid team (t-statistic)
6.1	Mean productivity[a] (ideas per group per 30-min)	22	26.23***	28.45	11.82	16.64*** (5.12)
6.2	Mean quality[b]					
	Business value (1–10 scale)	8,950	22.50***	4.79	4.52	0.27*** (4.74)
	Purchase intent (1–10 scale)	18,841	71.35***	4.93	4.58	0.35*** (8.45)
6.3	Within-team variance[b]					
	Business value	8,950	2.34	6.42	6.63	−0.21 (−1.53)
	Purchase intent	18,841	2.41	8.23	8.06	0.17 (1.55)
6.5	Quality of top five generated ideas[b]					
	Business value	2,157	69.55***	6.03	5.18	0.85*** (8.34)
	Purchase intent	4,535	151.14***	6.20	5.30	0.90*** (12.29)
6.6	Quality of top five selected ideas[b]					
	Business value	5,720	2.95	4.63	4.77	−0.15 (−1.72)
	Purchase intent	11,841	24.91***	4.95	4.63	0.32*** (4.99)
7.1	Degree of buildup[b]	7,745	19.42***	2.20	2.41	−0.21*** (−4.41)

[a]The unit of analysis is *organizational unit*.
[b]The unit of analysis is *idea rating*.
[c]Results are reported from an ANOVA with random effects for raters and/or creators. Identical results are obtained when raters and/or creators are introduced as fixed effects.
[d]Least square means are the mean residuals after taking into account the other control variables.
***$p < 0.0001$.

number of ideas generated, (b) the *average quality ideas*, (c) the *variance in the quality of the ideas*, and (d) the *ability of the process to discern the best ideas*. We then report on comparisons of the best *generated* ideas and the best *selected* ideas.

All hypotheses related to idea quality are tested using both business value and purchase intent as measures of quality. Unless stated otherwise, we use an analysis of variance (ANOVA) of the judges' ratings given each idea. That is, each rating of an idea provided by an independent judge is the dependent variable for a separate observation. The explanatory variable is the *treatment* (team versus hybrid). We include controls for the four-person group of individuals generating the ideas (the "creator") and the rater who provided the rating. This is because there are substantial differences in ability across the groups, and because there are systemic differences in how the scales were used by different raters. We considered the rater and creator effects as both fixed effects and random effects. Our results are nearly identical in either case. Furthermore, a Hausman test verifies the appropriateness of the use of the random-effects estimators.[2]

6.1. Effect of Group Structure on the Number of Ideas Generated

Table 2, row 6.1, illustrates the results of an ANOVA of the productivity, or the *number of ideas generated* in the two treatments, given the same number of people working for the same amount of time. The value shown is the number of ideas generated by the four-person group in 30 minutes. We control for the effects of the sets of individuals generating ideas and consider two alternate specifications, one with the creators as a random effect and a repeated-measures analysis. Our results are almost identical in the different specifications. We find that productivity is very

[2] The Hausman test compares the estimates from the more efficient random-effects model against the less efficient but consistent fixed-effects model to make sure that the more efficient random-effects model also gives consistent results.

Table 3 Rank Correlation Between Self-Assigned Ranks and True Ranks

	Rank correlation for business value			Rank correlation for purchase intent		
Treatment	Spearman	Kendall tau *b*	Hoeffding dependence	Spearman	Kendall tau *b*	Hoeffding dependence
Hybrid	0.162** (0.013)	0.121** (0.012)	0.005** (0.035)	0.182*** (0.005)	0.137*** (0.005)	0.008*** (0.009)
Team	0.082 (0.580)	0.051 (0.648)	−0.008 (0.865)	0.095 (0.519)	0.062 (0.577)	−0.007 (0.808)

**$p < 0.05$; ***$p < 0.01$.

different across different treatments; the hybrid process generates about three times more ideas than the team process (significant at the 0.01% level). This result is in line with our predictions and also in line with existing research. To the best of our knowledge, we are the first to verify these results statistically in a within-subjects design that controls for individual effects.

6.2. Effect of Group Structure on the Average Quality of Ideas

Table 2, row 6.2, shows the comparison of the *average quality* between the two different treatments. We evaluate and test the statistical significance of the difference in quality and conclude that the *hybrid process generates ideas of better average quality*. The quality advantage of the hybrid treatment is 0.25 units of business value and 0.35 units of purchase intent (significant at the 0.01% level for both business value and purchase intent). Although the magnitude of this difference may not appear large relative to the 10-point scale, a difference this large roughly translates to 30 points in percentile ranking (after controlling for fixed effects); in other words, this can be the difference between the 1st and the 30th idea in a pool of 100 ideas.

6.3. Effect of Group Structure on the Within-Group Variance in Idea Quality

Recall that we are interested in the variance in the quality of ideas *within a particular group*. We define this variance measure as the squared difference of the rating received by an idea and the average rating received by all ideas generated by the group in the specific treatment. We then conduct an ANOVA for this variable. The results are reported in Table 2, row 6.3. We do not find strong evidence for a difference between the team process and hybrid process as far as the variance of idea quality is concerned. In §7.4, we dig deeper into this finding by examining the extent of interactive buildup as a moderating variable.

6.4. Effect of Group Structure on the Ability to Discern Quality

We measure the *ability to discern quality* as the rank correlation between the preference ordering implied by the independent judges' ratings and the self-evaluation by the idea-generating group. As with all previous results, we provide this analysis for both business-value ratings and the purchase-intent ratings. The results are provided in Table 3. Note that the absolute value of the correlation for both team and hybrid groups is relatively low, in the best case, less than 0.2. This suggests that, irrespective of group structure, the ability of idea generators to evaluate their own ideas is extremely limited, and is perhaps compromised by their involvement in the idea generation step. Second, the hybrid structure has a significantly higher *ability to discern quality* than the team structure, supporting our theoretical prediction. In a further analysis, we compared the self-evaluation provided in the individual phase of the hybrid treatment to the independent judges' quality ratings, and found that these individual ratings are better predictors of "true quality" than are either of the group evaluations, lending further support to the idea that some aspect of the group interaction leads to poor assessments of quality.

6.5. Quality of the Best Generated Idea

Given the above results that the hybrid structure has a higher mean quality, higher productivity, and variance that is not systematically different, we expect that the quality of the best generated ideas to be higher for the hybrid structure. To test this prediction, we conduct an ANOVA of the ratings received by the top five ideas generated by each group. Table 2, row 6.5, shows the results from the comparison of the average quality of top five ideas in different treatments. We also test alternate versions, with the top three, four, and six ideas. In each of these cases our results provide similar support. As before, we include controls for the group of individuals generating the ideas, the rater who provided the rating, and the challenge to which the idea is addressed.

The ANOVA shows that the team and hybrid structure are different in the quality of the top five ideas. In particular, we evaluate and test the statistical significance of this difference and find that, as predicted, the top five ideas generated in the *hybrid process are of* *better quality* than those from the team process. Interestingly, the difference between the team and hybrid in terms of the quality of *best* ideas is much higher than the difference in *mean* quality of ideas. This is an illustration of how in an innovation setting, examining only mean quality as opposed to the quality of the best ideas is likely to underestimate the benefits of the hybrid approach.

6.6. Quality of the Best Selected Ideas

The creative problem-solving process includes both idea generation and idea selection. Unlike the analysis of §6.5, in this section, we will include the impact of idea selection. To do so, we compare the quality of the top five *selected* ideas between the hybrid and team structures. To test this hypothesis, we conduct an ANOVA on the independently determined quality ratings for the top five selected ideas. Table 2, row 6.6, shows us the results from the comparison of the average quality of top five selected ideas in different treatments. Using the purchase-intent quality metric, we conclude that the hybrid process results in higher quality for the top five selected ideas. For the business-value quality metric, we do not observe a significant difference in the quality of the top five ideas selected by each of the two structures. This result and the result on the comparison of the best generated idea in the preceding section suggest that the hybrid process may generate better ideas, but that because of the noisy selection process, its relative advantage is much diminished, to the point of becoming statistically insignificant for one of our quality metrics.

7. The Role of Buildup of Ideas

The results of the previous sections show that the hybrid process generates better ideas. Thus, the interactive buildup effect theorized for teams must be weak, at least when compared to the other effects in our theoretical framework. Our experimental design allows us to measure the extent of buildup at the idea level. In particular, recall that we coded the content of all ideas and computed the content similarities between consecutive ideas, which gives us a metric of the extent of buildup for these ideas.

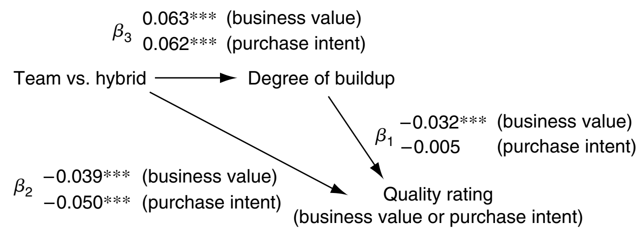
In this section, we first test whether individuals working in teams are more likely to build on ideas than individuals working in the hybrid structure. Next, we investigate the impact of this buildup on the underlying variables in our statistical framework.

7.1. More Buildup in Teams?

The existing literature has argued that teams are more likely to build on ideas. Recall that the buildup score is a measure of the extent to which an idea is similar to the previous idea. Table 2, row 7.1, shows the results from an ANOVA of the buildup scores of ideas. The results support the observation in the literature that ideas generated in teams are more likely to build on previous ideas.

Figure 4 **Two-Stage Least-Squares Model and Coefficient Estimates for Effect of Buildup on Idea Quality**

$$Buildup_i = \alpha' + \beta_3\ Team\ vs.\ hybrid_i$$
$$Quality\ rating_{ij} = \alpha + \beta_1\ Buildup_i + \beta_2\ Team\ vs.\ hybrid_i + \beta_4\ Rater_j$$



Notes. Results are presented with standardized coefficients obtained from a maximum likelihood estimation of the 2SLS model. The subscript i is an index for the idea, and j indexes the rater. Team versus hybrid effect is shown using an indicator for the team process. Business value: $N = 7{,}623$. Purchase intent: $N = 16{,}047$.
$^{***}p < 0.01$.

7.2. Impact of Buildup on Mean Quality of Ideas Generated

To investigate the impact of buildup on mean quality, we cannot conduct a direct regression (nor ANOVA) of quality on buildup. Such an approach would lead to incorrect estimates because both quality and buildup are influenced by an omitted variable in this regression, *organizational structure* (team versus hybrid). In other words, the error term in such a direct regression will include the effect of the process, and this would be correlated with the dependent variable. Thus, to test this effect we employ a two-stage least-squares (2SLS) procedure. The estimated equations, the proposed path model, and the standardized results from this model are illustrated in Figure 4.

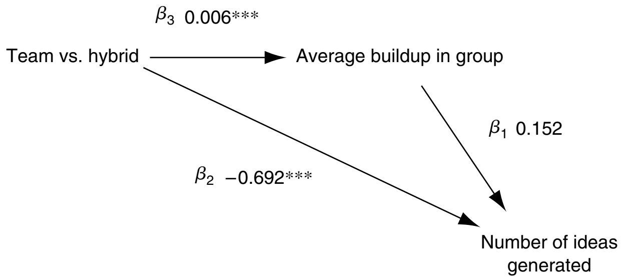
The results of our path analysis confirm the previously observed direct effect of organizational process on the quality and the extent of buildup. However, we find no support for the often-cited effect of buildup on improving the quality of ideas. In fact, in one of our models, we find the *reverse* effect: because of increased buildup, we observe that the mean idea quality actually decreases. This suggests that although teams indeed build on each other's ideas, this does not improve the quality of the ideas.

7.3. Impact of Buildup on Number of Ideas Generated

Next, we analyze the impact of buildup on the number of ideas generated. Although we see that buildup does not influence the quality of the ideas generated, it could still prove beneficial by expanding the number of opportunities that a group identifies. To test this effect, we compute the average buildup in

Figure 5 Two-Stage Least-Squares Model and Coefficient Estimates for Effect of Buildup on the Number of Ideas Generated

$$\text{Average-buildup-in-group}_{kl} = \alpha' + \beta_3\ \text{Team vs. hybrid}_{kl}$$
$$\text{N-Ideas}_{kl} = \alpha + \beta_1\ \text{Average-buildup}_{kl} + \beta_2\ \text{Team vs. hybrid}_{kl}$$



Notes. Results are presented with standardized coefficients obtained from a maximum likelihood estimation of the 2SLS model. The subscript k is an index for the group, and l is an index for the organizational process or treatment. Team versus hybrid effect is shown using an indicator for the team process. $N = 22$.

$^{***}p < 0.01$.

a group (following the team or hybrid process) and examine its impact on the number of ideas generated by the group. We follow the same empirical methodology as in the previous section. The estimated equations, the proposed path model, and the standardized results from this model are illustrated in Figure 5.

Again, although there is more buildup in teams, this buildup has no impact on increasing the number of ideas generated. This further illustrates that the beneficial consequences of buildup may have been over estimated in the prior literature. One explanation for this result is that the countervailing effect of production blocking is so strong that it completely dominates the potential productivity gain from buildup.

7.4. Impact of Buildup on the Variance of the Quality Distribution

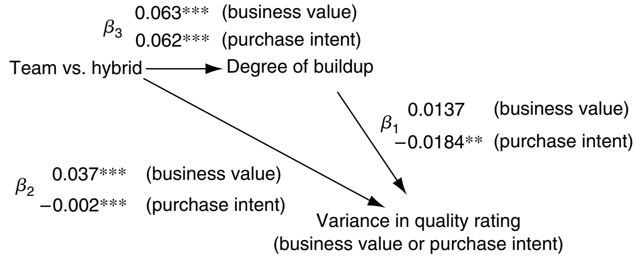
Interactive buildup is modeled as similarity in the *content* of consecutive ideas. This content similarity is higher in teams and, arguably, this should reduce the variance in quality of the generated ideas. Next, we examine the impact of the ideation process (team or hybrid) on the variance in quality, with the *extent of interactive buildup* as a moderating or path variable. Empirically, this is the same model as in the above two sections. The estimated equations, the proposed path model, and the standardized results from this model are illustrated in Figure 6.

As in the analysis of the direct impact of group structure on *variance in quality* (§6.3), we find that results vary by quality metric. Using purchase intent, we find that higher buildup in teams leads to lower variance; in fact, the direct impact observed before is fully moderated by the buildup. This suggests that teams indeed exhibit lower variance in part because of interactive buildup.

The analysis using business value suggests that teams exhibit higher variance (as in the analysis of

Figure 6 Two-Stage Least-Squares Model and Coefficient Estimates for Effect of Buildup on Variance in Idea Quality

$$\text{Buildup}_i = \alpha' + \beta_3\ \text{Team vs. hybrid}_i$$
$$\text{Variance-in-quality}_{ij} = \alpha + \beta_1\ \text{Buildup}_i + \beta_2\ \text{Team vs. hybrid}_i + \beta 4\ \text{Rater}_j$$



Notes. Results are presented with standardized coefficients obtained from a maximum likelihood estimation of the 2SLS model. The subscript i is an index for the idea, and j indexes the rater. Team versus hybrid effect is shown using an indicator for the team process. Business value: $N = 7{,}623$. Purchase intent: $N = 16{,}047$.

$^{***}p < 0.01$.

the direct impact) but this result is not explained by interactive buildup. In fact, interactive buildup has no impact on the variance in quality based on the business-value metric.

8. Conclusions and Managerial Implications

In this study, we compare the effectiveness of two group structures for solving problems that require creative idea generation followed by selection. First, the group of individuals can work as a team in the same time and space. Alternately, in a hybrid structure, the group works individually for some fraction of the time and then works together. In contrast with the vast body of existing literature that uses the number of ideas or the average quality of ideas as a performance metric, we use the quality of the best ideas as our metric. We find strong support that the best ideas generated by the hybrid structure are better than the best ideas generated by a team structure. This result is driven by the fact that the hybrid structure generates about three times as many ideas per unit of time, and that these ideas are of significantly higher quality on average. The hybrid structure is also better at identifying the best ideas; however, we find that both approaches do poorly in absolute terms in selecting the best ideas.

Our findings also shed light on one of the longstanding arguments for teams, the benefits of interactive buildup. We show that the suggested advantage of team-based brainstorming is not supported by experimental evidence. On average, ideas that build on other ideas are not statistically better than any random idea. This has significant managerial implications: if the interactive buildup is not leading to better ideas, an organization might be better off relying on asynchronous idea generation by individuals using, for example, Web-based idea management systems, as this would ease other organizational constraints such as conflicting schedules of team members and travel requirements.

We also investigated the impact of the two structures on the variance in the quality of different ideas. Here our results are inconclusive. While using purchase intent as a measure of quality, we find that teams exhibit lower variance in quality on account of the interactive buildup, but the results do not hold for the business-value metric. In fact, we find that the variance in the business value of ideas is higher for the team process. This issue merits further study.

As with any experimental study, the results may not hold in all settings. Our results on the quality of the best ideas depend not just on the directional comparisons between the two structures, but also on the magnitude of these differences. Although our experiment was set up to closely match problems in real-world settings, the subjects' limited time, resources, and prior exposure to the problem-solving context limit our ability to perfectly mimic a real situation. Furthermore, although the subjects were trained in ideation techniques and knew each other somewhat, they were not placed in teams that had developed a deep working relationship.

In all our results, we found that differences in performance *across individuals* are large and highly significant. These large performance differences suggest an interesting opportunity for future research. It would be interesting to examine whether these differences are persistent. If they are, an optimal process may be to first screen the pool of individuals for the highest performers and then employ only those individuals in subsequent idea generation efforts. However, the dynamics of the interaction between these high-performing individuals may differ significantly from the existing evidence and need to be explored in further experiments.

Acknowledgments
The authors are grateful to Christoph Loch, Kamalini Ramdas, Lee Fleming, the associate editor, and the anonymous reviewers for many helpful comments.

Appendix

A.1. Formal Statement of Theorems and Proofs

Theorem 1 (Effect of Number of Ideas). $\mathbf{E}[M_n] \leq \mathbf{E}[M_{n+1}]$.

Proof. Note that $\Pr[M_n \leq z] = \prod_{i=1}^{n} \Pr[X_i \leq z]$. Thus, the cumulative distribution function of the distribution of M_n, $G(z)$ is $F^n(z)$.

$$\mathbf{E}[M_n] = \int_0^\infty z g(z)\,dz = \int_0^\infty (1 - G(z))\,dz = \int_0^\infty (1 - F^n(z))\,dz.$$

Because $F(z) \leq 1$, $F^{n+1} \leq F^n$ and $1 - F^{n+1} \geq 1 - F^n$. The result now follows. □

Lemma 1. *If the quality of ideas generated follows a generalized extreme value (GEV) distribution* (Coles 2001) *with parameters* (μ, σ, ξ), *the quality of the best of n ideas also follows a GEV distribution with parameters*

$$\mu' = \mu + \frac{\sigma}{\xi}(n^\xi - 1), \quad \sigma' = \sigma n^\xi, \quad \text{and} \quad \xi' = \xi.$$

Proof. The result follows from substituting the cumulative distribution functions and reparameterizing. □

A similar result has been shown by both Dahan and Mendelson (2001) and Kavadias and Sommer (2009). Whereas Dahan and Mendelson (2001) work with the three different subfamilies of the generalized extreme value distributions, we present our result within the unifying framework of the GEV distribution. Kavadias and Sommer (2009) present this result for the Gumbel distribution. Also, note that the GEV distribution represents a fairly flexible family of distributions that can capture a wide variety of censored data. Because idea generation often involves some internal censoring by the ideator, this family is an ideal candidate for capturing idea quality. Furthermore, from data collected under a variety of ideation settings in real organizations, we find this family to be a reasonable fit.

Theorem 2 (Effect of the Mean of the Idea Quality Distribution). *Consider two ideation processes with GEV quality distributions with different means. All other central moments of the distributions are identical. The processes generate the same number of ideas. The expected quality of the best idea from the ideation process with the higher mean is higher.*

Proof. Because all moments besides the mean are identical for the two distributions, only the location parameter of the two quality distributions μ can be different, say $\mu_1 > \mu_2$. From Lemma 1, the best idea from each of the ideation processes will also be distributed GEV, with all parameters identical except the location parameters $\mu_1' > \mu_2'$. The mean of GEV distribution increases in the location parameter, and the result now follows. □

This result shows that all else being equal, the quality of the best idea from a process with a higher average quality is higher.

Theorem 3 (Effect of the Variance of the Idea Quality Distribution). *Consider two ideation processes with GEV quality distributions with different variance. All other central moments of the distributions are identical. The processes generate the same number of ideas. The expected quality of the best idea from the ideation process with the higher variance is better iff* $\Gamma(1 - \xi) > 0$.

Proof. Consider two GEV distributions, (μ_1, σ_1, ξ_1) and (μ_2, σ_2, ξ_2). The conditions on the central moments of the two distributions imply that $\xi_1 = \xi_2 = \xi$. $\sigma_1 \neq \sigma_2$; say $\sigma_1 > \sigma_2$ and $\mu_1 - \mu_2(\sigma_1 - \sigma_2)(1 - \Gamma(1 - \xi))/\xi$. From Lemma 1, the quality of the best idea from each of the ideation processes will also be distributed GEV, with parameters $(\mu_1 + (\sigma_1/\xi)(n^\xi - 1),\ \sigma_1 n^\xi, \xi)$ and $(\mu_2 + (\sigma_2/\xi)(\eta^\xi - 1),\ \sigma_2 n^\xi, \xi)$ and means $\mu_1 + (\sigma_1/\xi)(n^\xi \Gamma(1 - n) - 1)$ and $\mu_2 + (\sigma_2/\xi)(n^\xi \Gamma(1 - \xi) - 1)$; Γ is the gamma function. The result will hold if $(n^\xi - 1)\Gamma(1 - \xi)/\xi < 0$. Now note $n > 1 \Rightarrow (n^\xi - 1)/\xi > 0$. The result follows. □

Corollary 1. *Consider two ideation processes with Gumbel quality distributions with different variances. All other moments of the distributions are identical. The processes generate the same number of ideas. The expected quality of the best idea from the ideation process with the higher variance is better.*

Proof. The Gumbel distribution belongs to the GEV family with $\xi \to 0$. The result follows from an application of the above theorem and assuming $n > 1$. □

Theorem 4(a) (Coles 2001). *If there exist sequences of constants $\{a_n, b_n\}$ such that*

$$\Pr\{M_n^* \le z\} \to G(z) \quad \text{as } n \to \infty$$

for a nondegenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

defined on $\{z\colon 1+\xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$.

(b) *Given $\{Z_1, Z_2, \ldots, Z_m\}$, m observations of M_n, the parameters of $G(z)$ can be estimated as the* arg max *of the* log-*likelihood function*

$$l(\mu,\sigma,\xi) = -m\log\sigma\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-1/\xi},$$

provided that $1+\xi((z_1-\mu)/\sigma) > 0$, for $i = 1, \ldots, m$. As always, with maximum likelihood estimation, the parameter estimates are asymptotically normally, and approximate confidence intervals can be constructed using the observed information matrix.[3]

Proof. (a) The result is well known, and we refer the reader to Coles (2001) for an outline of the proof and to the references therein for a more technical version of the proof.

(b) Under the assumption that $\{Z_1, Z_2, \ldots, Z_m\}$ are independent variables having the GEV distribution, the above log likelihood follows from simple computation and absorbing the constants within the estimated parameters in the usual way. □

[3] A potential difficulty with the use of maximum likelihood methods for the GEV concerns the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. These conditions are not satisfied by the GEV model because the end points of the GEV distribution are functions of the parameter values: $\mu - \sigma/\xi$ is an upper end point of the distribution when $\xi < 0$, and a lower end point when $\xi > 0$. Smith (1985) considers this problem in detail and find that for $\xi > -1$, the estimators are generally obtainable and often have the usual asymptotic properties.

A.2. Subsample of Ideas Generated

Title	Descriptions	Mean rating
Mouth guard holder	A small, convenient, removable pocket that can be used to hold a mouth guard in between uses on the field	4.1
Odor reducing trash can	A trash can that reduces odor of garbage inside it	6.5
Water bottle with filter system	A water bottle with a built-in filtration system	5.9
Transforma-racquet	An athletic racquet that can be adjusted to accommodate any racquet sport	4.2
Waterproof reading system	A system for reading in the shower	3.2
Disposable desktop cover	This product is meant to be placed over a clean desktop; as clutter builds up, just fold up the cover and pull the draw string to trash the collected garbage	3.5
Toilet table	A foldable table that attaches to the toilet so you can read, eat, or do work while going to the bathroom	3.8
Coffee table with built-in remote	A coffee table that has a TV remote built into it so that you do not have to move far to change channels, but at the same time you do not have to search for a lost remote	3.7
Ball bag	A ball that functions as a bag until it is time to use it; when the ball is emptied, it then turns into a ball to use	3.4
Motion detection light	A light that detects that someone is trying to turn it on; when it senses motion at close proximity to the senor, it will automatically turn on or off	3.6
Hair collecting comb	A comb that collects stray hairs and makes them easy to dispose	5.3
Chore meter	A system that logs who did what chores at a certain time to establish who is not carrying their load	3.9
Noise reduction pad	A pad that is placed on the floor of a dorm room to reduce the level of noise heard by the room below; designed for students that work out in their rooms	5.5

A.3. Idea Categorization Scales

A.3.1. Challenge 1: Sports and Recreation. Ideas generated in Challenge 1 (sports and fitness products) were classified along the dimensions of "type of product," "principal sporting activity," and "key benefit proposition" in the following categories:

Type of product	Principal sporting activity	Key benefit proposition
Bag	Basketball	Convenience
Bottle	Bicycling	Hi-tech
Clothing	Field sports	Multipurpose
Gear and equipment	Golf	Hygiene
Food and drink	Gym/strength/fitness	Portability
Locks/security	Tennis and racquet sports	Customization/ personalization
Music/ entertainment	Running	Weather protection
Footwear	Swimming	Health
Information systems	Winter sports	Style
Watch	Not specific to activity	Reminder
	Other sport/activity	Eco-friendly

A.3.2. Challenge 2: Dorm and Apartment. Ideas generated in Challenge 2 (dorm and apartment) were classified along the dimensions of "type of product," "primary room or location," and "key benefit proposition" in the following categories:

Type of product	Primary room or location	Key benefit proposition
Apparel/accessories	Any	Convenience
Cleaning	Kitchen	Portability
Clocks, watches, alarms	Living	Multipurpose
Electronics/TV/ audio/computing	Bathroom	Hygiene
Food, cooking, and eating	Bedroom	Customization/ personalization
Furniture and décor	Study/office/ desk area	Automation
Heating, ventilation, air conditioning	Walls	Hi-tech
Lighting	Garden/outdoors	Style
Personal care and health	Closet	Disposable
Power management and electricity		Reminder
Security		Safety
Storage		Value/low cost

References

Amabile, T. M. 1996. *Creativity in Context*. Westview Press, Boulder, CO.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational Psych. Measurement* **20**(1) 37–46.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag, London.

Dahan, E., H. Mendelson. 2001. An extreme value model of concept testing. *Management Sci.* **47**(1) 102–116.

Diehl, M., W. Stroebe. 1987. Productivity loss in idea-generating groups: Toward the solution of a riddle. *J. Personality Soc. Psych.* **53**(3) 497–509.

Diehl, M., W. Stroebe. 1991. Productivity loss in idea-generating groups—Tracking down the blocking effect. *J. Personality Soc. Psych.* **61**(3) 392–403.

Gwet, K. 2002. *Handbook of Inter-Rater Reliability*. STATAXIS Publishing, Gaithersburg, MD.

Jamieson, L., F. Bass. 1989. Adjusting stated purchase intentions measures to predict trial purchase of new products. *J. Marketing Res.* **26**(3) 336–345.

Janis, I. L. 1982. *Groupthink*. Houghton Mifflin, Boston.

Kavadias, S., S. C. Sommer. 2009. The effects of problem structure and team diversity on brainstorming effectiveness. *Management Sci.* **55**(12) 1899–1913.

Kornish, L. K., K. T. Ulrich. 2009. Characterizing opportunity spaces in innovation: Evidence from large samples of ideas in five domains. Working paper, http://ssrn.com/abstract=1503103.

Mullen, B., C. Johnson, E. Salas. 1991. Productivity loss in brainstorming groups: A meta-analytic integration. *Basic Appl. Soc. Psych.* **12**(1) 3–23.

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

Paulus, P. B., V. Brown et al. 1996. Group creativity. R. E. Pursuer, A. Montuori, eds. *Social Creativity in Organizations*. Hampton, Creskill, NJ.

Pugh, S. 1981. Concept selection: A method that works. *Proc. Internat. Conf. on Engrg. Design, Rome*, 9–13.

Robbins, S. P., T. A. Judge. 2006. *Organizational Behavior*. Prentice Hall, Upper Saddle River, NJ.

Singh, J., L. Fleming. 2010. Lone inventors as sources of breakthroughs: Myth or reality? *Management Sci.* **56**(1) 41–56.

Stroebe, W., M. Diehl. 1994. Why are groups less effective than their members: On productivity losses in idea generation groups. *Eur. Rev. Soc. Psych.* **5** 271–303.

Sutton, R. I., A. Hargadon. 1996. Brainstorming groups in context: Effectiveness in a product design firm. *Admin. Sci. Quart.* **41**(4) 685–718.

Taylor, A., H. R. Greve. 2006. Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Acad. Management J.* **49**(4) 723–740.

Terwiesch, C., C. H. Loch. 2004. Collaborative prototyping and the pricing of custom-designed products. *Management Sci.* **50**(2) 145–158.

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

Ulrich, K. T., S. Eppinger. 2007. *Product Design and Development*. McGraw-Hill Higher Education, New York.

Van de Ven, A. H. 1986. Central problems in the management of innovation. *Management Sci.* **32**(5) 590–607.

Wageman, R. 1995. Interdependence and group effectiveness. *Admin. Sci. Quart.* **40**(1) 145–180.