Organizational vs. Crowd Selection: Implications for Exploration and Exploitation

Berk Can Deniz[1] and Jesper B. Sørensen[1]

[1]Stanford University, Graduate School of Business

September 14 2019

Abstract

This article considers how the interdependence between the variation and selection stages of the innovation process affects the variety of ideas generated. Most organizations rely on internal selection mechanisms, whereby creators submit their ideas to managers or experts within their firm for approval. Recent years have seen a growth in the uses of crowd-based selection mechanisms, whereby external audiences choose among ideas. While prior work has compared crowds and expert in terms of which kinds of ideas they select, we examine how these alternative selection mechanisms might influence the idea-generation process. We argue that internal selection generally lowers exploration by reducing the variation in ideas because creators with a clear conception of selection criteria constrain their search for ideas. We use two separate innovation tournaments to compare the effects of selection mechanisms on the variety of ideas generated. The findings are consistent with the claim that internal selection reduces variation. The results have implications for both the theory and practice of organizational innovation.

1 Introduction

It has long been accepted in the literature on organizational learning and adaptation that firms facing environmental change benefit from being able to generate a variety of ideas and innovations internally. The variety of ideas produced by a firm can be viewed as a distinct dimension of innovation and creativity, where an organization is able to develop ideas in a number of different domains of knowledge or product categories. The ability to generate a variety of ideas is central to organizational exploration (March 1991; Levinthal 1997). The greater the variety of ideas, the higher the probability of successful discovery of new market opportunities and the less likely the firm is to land in a competence trap due to an over-emphasis on exploitation. Creativity without a variety of ideas, by contrast, may be valuable but this value will come from the exploitation of established knowledge and competencies.

A substantial literature has focused on how organizational characteristics, policies, and practices shape the balance between exploration and exploitation (Sorensen and Stuart 2000; Benner and Tushman 2002; Ederer and Manso 2013). This literature has highlighted that the balance between exploration and exploitation arises through the two channels: organizational variation processes and organizational selection and retention processes. Organizational variation processes refer to policies and practices that cause the organization to sample new ideas and experiment with new possibilities, while selection and retention processes are devoted to filtering out unwanted solutions and transferring preferred ideas and solutions throughout the organization. In this view, a firm can increase exploration by devoting resources to variety-generating activities, or by shifting the nature or extent of its selection and retention processes.

Thus a firm can attempt to increase the degree of exploration by devoting more resources to idea-generating activities, whether those be external search or internal experimentation and recombination. 3M Corporation's "bootlegging" policy, as well as Google's "20% time", where engineers were allowed to use company time and resources to pursue their ideas, are classic examples. Investing resources in variety-generating processes is a way of turning up the degree of exploration. The primary way in which firms might increase the degree of exploitation is through the use of organizational and managerial selection and retention mechanisms that are based on the firm's established base of knowledge. Benner and Tushman (2002) found, for example, that as firms increase their emphasis on process management, their innovations became more exploitative of the firm's existing knowledge. Similarly, firms in which there is a strong cultural consensus are biased toward exploitation, as the strong culture reinforces taken-for-granted ways of seeing things (Sorensen 2002).

A bias toward exploitation can arise, in short, either from insufficient investment in variety-generating activities or in the use of selection processes that implicitly or explicitly privilege existing knowledge or both. The literature on exploration and exploitation in firms has placed primary emphasis on the role of internal selection processes in generating a bias against exploration (O'Reilly and Tushman 2013). A major lesson from this literature is that much of what managers do, especially in the form of standard operating procedures and routines, creates a strong bias toward exploitation in established firms because these processes work to limit the variety of ideas. Even when firms invest heavily in generating variation, the ways in which internal processes treat ideas coming out of the variation stage may push the firm toward exploitation despite its professed interest in exploration. This is most obvious when the organization's selection criteria are explicitly based on alignment with the established strategy, but can also occur when these criteria are only implicit in the mental models of the decision-makers at the selection stage.

Much of the literature on exploration and exploitation are rooted in traditional models of formal organization, in which innovation and creativity are seen as largely intra-organizational phenomena. Yet transformations in management and technology (particularly the growth of the Internet) have made it easier for organizations to reach out to crowds. Many modern organizations frequently use the power of crowds both to generate and to select the ideas with the greatest potential (Felin, Lakhani, and Tushman 2017).

Much of the scholarly interest in crowd-based innovation has been in the potential for the crowd to increase the diversity of ideas at the variation stage by sourcing ideas outside the organization, and thereby increase the potential for exploratory learning. But crowds also increasingly play a role at the selection stage. Many companies now try to engage with markets as early as possible to collect feedback instead of relying on in-house managers and experts to evaluate ideas. Concepts such as experimentation, rapid prototyping, and beta testing are now discussed in the innovation and entrepreneurship world probably more than ever before. Many new management approaches, such as lean (Collis 2016), agile (Rigby, Sutherland, and Takeuchi 2016), and design thinking (Kolko 2015), emphasize the direct use of crowds and markets to evaluate the potential of ideas and products. A good example of this trend is Intuit, a company that made the decision to exchange "presentations to managers" with "experiments with customers" and experienced a remarkable positive impact in terms of innovation (Martin 2011).

The appeal of crowd-based selection mechanisms relative to internal selection mechanisms is that they are less likely to mistakenly reject early-stage ideas that turn out to have strong potential (Mollick and Nanda 2016). This is because internal selection mechanisms generally rely on a narrower domain of knowledge, and hence are more likely to be biased toward exploitation. Firms are therefore often interested in using crowd-based selection mechanisms as a means of assessing ideas generated in-house as a way of avoiding false negatives in particular – rejecting ideas that look unattractive according to internal selection criteria but that the market recognizes as valuable.

Our interest in this paper is in a different potential implication of the contrast between internal and external crowd-based selection mechanisms, namely the consequences of the selection mechanism for the diversity of ideas generated at the variation stage. Our expectation, detailed below, is that relying on external crowd-based selection mechanisms is likely to increase the variety of ideas generated at the variation stage, independent of any effects of the selection mechanism on the choice among ideas emerging from the variation stage. The idea here is simple: a creator is more likely to constrain their search for ideas, and hence generate less variety, to the extent that they have a clear conception of the criteria that are likely to be implemented at the selection stage. To the extent that creators are motivated to have their ideas accepted, they have incentives and tendencies to make inferences about their evaluators' tastes and then generate ideas according to these inferences. Formulating a clear conception of the evaluator's taste function is easier to the extent that the creator has some knowledge of the evaluator's identity, and to the extent that there are fewer evaluators. As a result, the variety of ideas will be lower when internal selection mechanisms are used rather than external selection mechanisms.

We test our theory by running two online innovation tournaments, where subjects are asked to generate ideas to be entered into an innovation tournament. Innovation tournaments are an established way of structuring innovation processes within firms (Terwiesch and Ulrich 2009). Here we use a mock innovation tournament to induce participants to propose innovative products and product applications in the hopes of winning a prize. The experimental conditions implement the contrast between internal and external selection mechanisms by manipulating the information provided to the subjects about the nature of the selection stage, specifically the identity and number of evaluators. The ideas generated by the subjects are then hand-coded in order to assess the variety of ideas generated. The findings suggest that creators who think they will be evaluated by a crowd generate a greater variety of ideas.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature and outlines the theoretical arguments. Section 3 describes the research design and measurement strategies. Section 4 and 5 presents the details, coding methods, dependent variables, and results of the first and second studies. Section 6 discusses the potential limitations and section 7 discusses the implications of the findings.

2 Theory

The dynamics of organizational idea generation and exploration have been extensively studied by scholars in various fields. Scholars who approached variation from macro perspectives analyzed the impact of organizations' age (Sorensen and Stuart 2000), organizational structure (Singh and Fleming 2010), and networks (Burt 2004), while scholars who analyzed variation from micro perspectives found evidence for team processes (Girotra, Terwiesch, and Ulrich 2010), individual characteristics (Anderson, Potočnik, and Zhou 2014), and motivation (Amabile and Pratt 2016). The interaction of the variation and selection stages of the idea generation and learning process has received relatively limited attention, however (Levinthal 2007; Knudsen and Levinthal 2007; Berg 2016; Girotra, Terwiesch, and Ulrich 2010).

In the literature on organizational learning, variation and selection are generally treated as sequentially interdependent stages. The (perhaps implicit) assumption is that the variation process is blind to the method of selection. This assumption is inherited from the origins of the variation-selection-retention model in biological evolutionary theories dating back to Darwin. In evolutionary biology, genetic mutations (variation) are blind to natural selection (selection). In other words, there is no mechanism through which an organism can change or optimize its genetic mutations in order to survive natural selection.

In the realm of creativity and innovation, the notion of blind variation may be appropriate when applied to the individual psychology of creativity, where variation and selection are both viewed as intra-individual processes (although even here there is substantial debate as to the extent to which variation is blind) (Simonton 2011). Similarly, it may be appropriate when applied to purely market-based innovation. Yet a large share of innovative activity takes place within organizations, and the creation of new ideas is a central concern of most organizations.

The assumption that variation is blind to the selection process breaks down in many if not all organizational contexts. Most organizations implement a division of labor between variation and selection, with the different stages carried out by different actors (Mollick 2012; Berg 2016). Most commonly, variation and selection are both performed in-house, although the emergence and growth of crowd-based innovation techniques have increased the ease with which firms can combine internal and external crowd processes.

One consequence of the division of labor between variation and selection roles is that generating ideas that survive a selection process is often associated with financial rewards, career progress, or higher status, whereas ideas that fail the selection stage might result in the opposite. Some organizations try to counteract fear of failure by celebrating or rewarding "well-intentioned" failure; the existence of such efforts is evidence of the nature of the concern that potential innovators have. Even in the absence of such career concerns, a creator's sense of self-worth and identity may be tied to how her ideas are received by an audience.

We can, therefore, expect creators to change or optimize their behavior according to the selection method they face. Indeed, an extensive and insightful literature on creativity in organizations has examined this issue. Work in this tradition has provided insights into how creativity or the level of novelty depends on whether creators are provided with intrinsic or extrinsic motivation (Amabile, Conti, Coon, Lazenby, and Herron 1996), on the specific incentive structures used (Ederer and Manso 2013), the nature of managerial control being exercised (Zhou 2003) and managerial goal-setting (McGrath 2001).

Studies in this tradition share a focus on the organizational design and managerial implementation choices that affect the consequences of internal selection mechanisms for creativity. In this respect, they are primarily focused on how work environments can be optimized for employee creativity. Moreover, the literature is primarily concerned with the generation of novelty, with less concern for the variety of ideas *per se*. Thus this work provides limited guidance for thinking about how the contrast between internal and external selection mechanisms will affect the *variety* of ideas generated.

In our view, a critical difference between internal and external crowd selection mechanisms is in the ability of creators at the variation stage to form a clear picture of the audience for their ideas. For a creator who is concerned about the acceptance of their ideas, formulating a clear picture of the evaluator's preferences and tastes is a critical factor. Having a clear conception of the evaluator's taste allows the creator (perhaps subconsciously) to target the evaluator's preferences as a means of increasing the likelihood of approval. In the absence of such clarity, by contrast, the search for new ideas will be less constrained.

From the standpoint of the diversity of ideas likely to be generated by a creator, the implications are clear: Creators with a clear conception of the evaluator's taste should generate less diverse ideas, as their ideas will be targeted at their perception of the evaluator's preferences. We term this process "taste targeting." Our claim is that to the extent that individuals engage in taste targeting at the variation stage, the diversity of the ideas that are generated will be lower. We argue that taste targeting is more likely to occur under internal selection regimes than external crowd selection regimes.

Note that from the point of view of a person charged with creating ideas, the two types of selection mechanisms diverge along at least three relevant dimensions: the availability of information about the identity of the person(s) performing the selection; the number of evaluators; and the existence of a hierarchical relationship.

The first distinction is the availability of identity information about the evaluators. Organizational life typically induces employees to interact regularly. Even in the absence of direct interaction, gossip networks mean that creators are likely to be aware of the reputation of those individuals making selection decisions. As a result, they acquire information about such things as the evaluator's expertise, tastes, and biases. This constitutes a stark contrast with crowd selection, where creators often lack any information about the people judging their ideas. The presence of identity information makes it easier for a creator to form a clear picture of the evaluator's taste function and hence should lower the diversity of ideas produced.

The second distinction between internal and external selection lies in the number of evaluators. While internal selection process typically relies on evaluation by one or a small number of people, crowd selection, by definition, relies on a large number of evaluators. We argue that an increase in the number of evaluators changes the behavior of creators by making it harder for them to develop a clear conception of the preferences to be targeted. Indeed, knowing that one's ideas will be assessed by a large number of evaluators makes the potential diversity of tastes salient. By contrast, imagining a single evaluator may allow the creator to imagine a clear target, even in the absence of identity information. As a result, a larger number of evaluators should result in a greater diversity of ideas produced.

Unless evaluators are from a narrowly defined group, a large number of evaluators results in complicating taste inference. It is not even clear that a large group of people (such as a market or a crowd) would necessarily have any particular taste. From the point of view of creators, having a single evaluator makes it easier and advantageous to target the evaluator's particular taste.

[FIGURE 1 GOES HERE]

The third distinction between internal and external selection mechanisms is the presence of a hierarchical relationship in the former. Working with a manager means that the person who makes judgments about creators' ideas is the same person who has control over creators' careers, including in domains unrelated to the assessment of particular ideas. Crowds do not, obviously, exercise the same form of control. We expect that hierarchical relationships increase taste targeting (and lower the diversity of ideas), other things being equal, as they create a greater incentive to submit ideas that will be approved by the evaluator.

A visual representation of our argument can be found in Figure 1.

3 Research Design

We test our theoretical arguments through two online experiments, where subjects are asked to generate ideas to be entered into an innovation tournament. Before discussing the specific features of each experiment, we first discuss the overall research design shared by the two studies.

An online experiment is an appropriate research design for our purposes, even though it sacrifices some external validity. In our view, the control provided by the online experiment has substantial benefits relative to the use of observational data or even field experiments. Evidence suggests that experiments performed through crowd-sourced online platforms such as Mechanical Turk (used here) maintain strong internal validity relative to experiments performed on nationally representative samples (Weinberg, Freese, and McElhattan 2014).

Two challenges present themselves with respect to observational data. First, generating reliable measures of the selection procedures used by established organizations, along with detailed information on the variety of ideas generated by organizational members, is difficult at best. Second, even with such data, one would be concerned about the endogeneity of the variation observed in organizational selection mechanisms or the extent to which the firm's decision to adopt internal vs. external crowd selection is related to the variation of ideas.

A field experiment would address these concerns to a certain extent, assuming that one or more organizations were willing to randomly assign idea creators in their organization to different selection mechanisms. However, securing this cooperation is difficult, apart from the challenges involved in implementing a clean experimental design. A field experiment also raises concerns about the endogeneity of the firm's decision to participate in a field experiment; perhaps firms with difficulty generating a variety of ideas, or a greater perceived need to generate a variety of ideas, are more likely to participate. Moreover, the effects observed in a field experiment in established firms would likely be confounding the content of change (i.e., the different selection mechanisms) from the process of change (i.e., the move from one selection mechanism to another) (Barnett and Carroll 1995).

3.1 Experimental Manipulations

We turn now to the overall design of our experimental manipulations. In our theoretical argument, we suggested that internal and external crowd selection mechanisms could usefully be differentiated along three dimensions: the existence of hierarchical relations; the existence of identity information about evaluators; and the number of evaluators.

Ideally, our experiments would implement variation along with all three of these dimensions. However, it was difficult to manipulate the degree of hierarchy in the relationship between the evaluator and the experimental subject. Doing so would have required inducing subjects to believe that the (fictitious) evaluator had authority over them. The nature of the online innovation tournament and the use of online subjects made this difficult to do in a reliable and valid way.

Instead, we independently manipulate the provision of identity information and the number of evaluators. These dimensions are cross-classified in Table 1.

[TABLE 1 GOES HERE]

In Table 1, Cell A corresponds to a classic internal selection process, where there are few evaluators, and the creators know something about them. Traditional superior-subordinate relationships (with the subordinate the idea generator) would fall in Cell A. While our studies below do not try to mimic an authority relationship, we operationalize another important aspect of internal selection mechanisms by providing information on the domain expertise of the evaluator. Internal selection mechanisms where ideas are evaluated by a committee (such as many new product introduction processes) would also fall in Cell A.

External or crowd selection is represented in Table 1 by Cells B and C. Each cell represents a different form of crowd selection. In Cell B, there are many evaluators, but creators know something about the characteristics of the evaluators. What we have in mind here are cases where creators know that the evaluators belong to a certain category — for example, that they have certain expertise, or perhaps certain demographic characteristics. Cell C, which we might consider the "pure" crowd condition, is, by contrast, one in which creators know nothing about the characteristics of the evaluators.

Cell A corresponds to the condition where we expect to see the highest level of taste targeting, and therefore the least diversity of ideas generated. Cell B should have lower levels of taste targeting, and hence a greater diversity of ideas than Cell A, while Cell C should have the least taste targeting and the greatest diversity of ideas.

Note that the contrast between Cells B and C is informative about the effects of identity information.

In this paper, we do not provide a test for the empty cell in Table 1, corresponding to no identity information and a small number of evaluators. While this case seems a theoretical possibility, it would seem to be a rare empirical phenomenon.

3.2 Measurement of Dependent Variable

In each of the two studies below, experimental subjects generate multiple ideas for participation in the innovation tournament. Each of those ideas is coded along multiple dimensions, which differ between the two studies and are described below. Each idea can, therefore, be characterized as a vector in a multidimensional space.

To assess the variety of ideas, our dependent variable, we want to measure the distance between the idea vectors. Conceptually, two ideas are similar to the extent that they can be represented by the same vector, i.e., they are categorized in the same way along the different dimensions identified in the coding scheme. Therefore, a simple way to capture the distance between each pair of ideas is to count the number of non-overlapping categories. The greater the number of non-overlapping categories, the greater the difference between ideas, hence, the distance.

Counting the number of non-overlapping categories is the equivalent of using Manhattan Distance or Hamming Distance Measures. If $(\mathbf{p}, \mathbf{q})$ are vectors representing two different ideas, the distance measure is defined as the sum of the absolute value of the differences:

$$d(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} |p_i - q_i|$$

We calculate $d(\mathbf{p}, \mathbf{q})$ for each pairwise combination of ideas submitted by each participant. Distances between ideas are only calculated within-participant; distances to the ideas submitted by other participants are not of interest since our interest is in the variety of ideas at the participant level.

We adopt two different approaches to characterizing the overall variety of the ideas submitted by a participant. The first is simply the mean distance between all pairwise combinations of ideas submitted by an individual participant.

A disadvantage of the first measure is that it may over-weight a cluster of similar ideas and understate the range of ideas submitted by a given individual. For example, a participant may submit six very similar ideas, but then the seventh and eight ideas are very different from the previous ideas, and from each other. In this scenario, a measure using the mean would imply that this participant has a lower variety than that of a participant who only submitted the sixth, seventh, and eighth ideas.

To the extent that exploration is about the range of ideas, we therefore also use as an alternative measure the average distance between an individual's three most distant ideas. In other words, among all of the pairwise distances between ideas submitted by an individual, we retain the three highest values and then compute the mean of those values.

4 Study 1

4.1 Study 1 Design

Our goal in Study 1 was to provide an initial test of our ideas by implementing the strongest contrast between internal and external selection mechanisms. We do this by varying the availability of identity information and the number of evaluators simultaneously. The two conditions in Study 1 correspond to Cells A and C of Table 1, which we for simplicity refer to as the expert and anonymous crowd conditions, respectively.

Study 1 consists of an online innovation tournament (Terwiesch and Ulrich 2009) called "The Future of Drones". Subjects were informed that they were participating in a tournament organized by a (fictitious) innovation and design company that had been hired by a start-up in the drone industry. The goal of the tournament was to generate new ideas "to find new uses for drones in order to expand [the drone industry's] customer base" because (fictitious) market research had shown that most drones were used for photography.

Two hundred subjects were recruited from Mechanical Turk and were paid $2 for participation in the study. The design of the tournament included a $30 bonus payment, in addition to the participation payment, for each of three winners. The bonus intended to strengthen the incentive to generate good ideas. Subjects were encouraged to generate, in an 8-minute period, as many ideas as they could. They were told to provide two to three sentences describing each idea they submitted. Moreover, participants were told that their ideas would be anonymized and judged independently. Subjects were randomly assigned to one of two conditions, each of which provided different information about who would be evaluating the ideas submitted in the tournament.

[FIGURE 2 GOES HERE]

- Participants in the expert condition were informed that "An expert in the field will rate the ideas" and "The expert will declare three winners." They were then provided with a short description of the expert, which is reproduced in Figure 2, along with a photo purporting to be the expert (actually the first author).

- In the anonymous crowd condition, subjects were informed that "A large group of potential customers will rate the ideas" and "The three most popular ideas will be declared winners." They were then provided with the following information about the crowd: "The judges of this tournament are online shoppers. Using a popular online shopping website, the customers were selected at random to be our judges. There will be a total of 1000 judges."

After reading the description of the judging process, subjects were given eight minutes to generate as many ideas as possible. An example of the idea entry page is provided in Figure 3.

[FIGURE 3 GOES HERE]

An attention check at the end of the experiment asked participants about the judgment process of the tournament. Participants were excluded if they did not correctly recall how their ideas would be judged, choosing from a list of alternatives that included the true judgment process. Approximately one-quarter of participants failed this attention check.

Finally, during the idea coding process, the ideas that were deemed incomprehensible by the authors were disqualified. In total, 5 out of 830 ideas were disqualified. The number of subjects satisfying the attention check and the average number of ideas are reported in Table 2.

[TABLE 2 GOES HERE]

4.2 Study 1 Coding

The first step in measuring the variety of ideas generated by subjects in each condition was to code the ideas. In the absence of an established coding scheme for drone usage ideas, we developed a coding scheme inductively based on a random sample of ideas (sampled blind to experimental condition).

Ideas were coded by the first author along two dimensions, corresponding to a general syntax identified in the idea proposals. The syntax consists of 1) a function, and 2) an object of a function.

For each idea, two questions were asked: "what is the function of this drone?", and "what is the object of the function?". Every idea was coded into one-word answers to these questions.

A classification scheme for the two dimensions was created inductively. If for the function and object dimensions, the content of an idea could not be assigned to an existing category, a new category was created. The goal was to create categories general enough to compress unimportant variation. To illustrate, compare the following two ideas:

> *Use drones to deliver pizza from delivery places. This will save on manpower for the pizza shop and let them save on wages.*

and

> *Food delivery: A drone could take off from a restaurant of your choosing and fly to your house without getting stuck in traffic.*

Although the first idea does not use the word "food", instead of creating a new variable for pizza, we treated it as a type of food. As a result, both of these ideas were coded as delivery (function), and food (object).

The resulting coding scheme contained 30 function categories and 33 object categories. Coding within the function and object dimensions were not mutually exclusive; thus an idea might include multiple objects or multiple functions. The codebook can be found in the supplemental materials.

Based on this coding process, each idea can be represented as a vector with 63 binary entries (30 function + 33 object). These are sparse vectors since ideas are generally relatively focused; 93% of the idea vectors have two non-zero entries. These vectors were then used to compute the measures of the variety of ideas described above.

4.3 Study 1 Results

Table 3 presents summary statistics for the two measures of the variety of ideas. Figure 4 graphs the distributions of the two different versions of the dependent variable by experimental condition. This figure provides initial evidence to support the claim that external crowd selection mechanisms generate a greater variety of ideas at the variation stage.

[TABLE 3 GOES HERE]

[FIGURE 4 GOES HERE]

Table 4 presents OLS regressions of our measures of the variety of ideas on the two experimental conditions. The two columns in the table correspond to different calculations of the mean distance between the ideas submitted by a participant; in the first column, the mean is calculated across all ideas, while in the second it is only calculated across the three most distant pairs of ideas. The intercept in these regression models is suppressed; the coefficient for the expert condition represents how the variety of ideas in that condition departs from the anonymous crowd condition.

[TABLE 4 GOES HERE]

Both sets of estimates in Table 4 are consistent with the notion of taste targeting. Study participants who expect their ideas to be evaluated by an anonymous crowd generate a greater variety of ideas than study participants who expect their ideas to be evaluated by an expert.

The results are substantively large. When computed over all pairs of ideas, the variety of ideas in the expert condition is approximately 12% lower than in the anonymous crowd condition (-0.463/3.854). Similarly, the estimated coefficient for the expert condition in the first column of Table 4 implies a reduction in the variety of ideas of approximately two-thirds of a standard deviation (-0.463/0.670). When computed over the three most distant pairs of ideas, the reduction is slightly smaller at 0.57 standard deviations, but still a substantial effect.

To summarize, the two OLS regressions show that the evaluation by a small number of known individuals leads people to generate less variation in their ideas relative to the evaluation by a large number of unknown people. We conclude that Study 1 provides empirical evidence consistent with our theory of taste targeting and the benefits of crowd selection for variation and exploration.

5 Study 2

5.1 Study 2 Design

Similar to Study 1, Study 2 consists of an innovation tournament. A primary goal of Study 2 was to separate the effects of identity information from the effects of the number of evaluators. Recall that in Study 1, our experimental conditions varied identity information and the number of evaluators simultaneously. Thus the estimates from Study 1 do not shed light on whether the effects are due to the presence of identity information or the number of evaluators (see Table 1). To tackle this problem, we introduced a third treatment called "known crowd" to Study 2. This new treatment is characterized by a large number of evaluators with known characteristics; specifically, their expertise is manipulated to be the same as the single evaluator.

The experimental manipulations of the selection mechanism were as follows:

- Participants in the expert condition were informed that their ideas would be evaluated by "Riley Ferguson, a leading marketing executive in the sporting goods industry, with extensive experience in developing innovative sports products in different market segments. Ferguson worked as a marketing specialist for Nike for more than a decade and was subsequently recruited as an executive by other leading firms in the sporting goods industry, including Wilson, Babolat, and Adidas." The participants were then provided with the following information: "The idea that receives the highest rating from Riley Ferguson will win the tournament." This condition corresponds to cell A in Table 1. The name Riley Ferguson was chosen to be gender-neutral, to reduce concerns that the salient identity characteristic of the expert condition in Study 1 was due to gender. Also, unlike in Study 1, no photo of the expert was provided, to reduce concerns about the potential effects of race and ethnicity.

- In the known crowd condition, subjects were informed that their ideas would be evaluated by "a group of executives with extensive experience in developing innovative sports products. 100 executives were recruited from the membership of the National Sporting Goods Association, including executives from Nike, Wilson, Babolat, and Adidas." They were then provided with the following information: "Each executive will evaluate ideas independently, and their ratings will be averaged. The idea with the highest average rating will win the tournament." This condition corresponds to cell B in Table 1.

- In the anonymous crowd condition, subjects were informed that their ideas would be evaluated by "a large group of potential customers. 100 potential consumers were recruited by contacting adults on Facebook who live in the United States." The participants were then provided with the following information: "Each potential consumer will evaluate ideas independently, and their ratings will be averaged. The idea with the highest average rating will win the tournament." This condition corresponds to cell C in Table 1.

The second goal of Study 2 was to shed greater light on the channels through which the selection mechanisms exert their effects on the variety of ideas observed. In particular, our goal was to clarify whether the differences in idea variation due to the selection mechanisms arise when participants are generating ideas, or when they choose which among their ideas should be submitted for evaluation. In other words, subjects in different conditions generate the same variety of ideas but submit different sets of ideas for evaluation, or does the selection mechanism operate directly on the variety of ideas generated (Yuan and Zhou 2008). These two processes were confounded in Study 1.

Following Yuan and Zhou (2008), in Study 2, we split the tournament into two stages. In stage 1, participants were asked to write down as many ideas as they can but told that these ideas would not be evaluated; in stage 2, they were asked to submit three ideas, from among those generated in the first stage, for evaluation.

A final goal for Study 2 was to perform the innovation tournament in a more established product category and make it more comparable to prior research. We therefore followed Girotra et al. (2010) and fielded a tournament about sports products. Subjects were informed that they were participating in an innovation tournament organized by a fictitious sports company that was looking for new products to be sold in a sporting goods retailer.

We recruited three hundred subjects, one hundred for each condition described below, from Mechanical Turk and paid them $2 for their participation. The design included a $100 bonus payment for the best idea.

Roughly 30% of the sample failed the attention checks. To be certain about the data quality, we asked the research assistants to vote on ideas that should be disqualified as well. A more detailed explanation of the disqualification process can be found in Section 5.2.

Table 5 provides the descriptive statistics about Study 2.

[TABLE 5 GOES HERE]

5.2 Study 2 Coding

Unlike Study 1, which required us to inductively create a coding scheme for drones, we used an established coding framework for sports products. Sporting goods retailers need to categorize their products and care about the accuracy of their categorization framework to manage their product portfolio and improve the customer shopping experience. Because these frameworks for categorizing products are relatively well-established and consensual, we used the categorization framework of one prominent sporting goods retailer, Dick's Sporting Goods, as the basis for our coding scheme to apply to the ideas in our tournament.

Using a modified version of the categorization scheme as our codebook, the ideas in the tournament were coded along two dimensions: 1) sports activity, which consists of 26 categories, and 2) product type, which consists of 25 categories. We created a codebook that contains category names, definitions, examples, and links to the company website for coders to reference if needed. The codebook can be found in the supplemental materials.

We hired three research assistants to code the ideas from the tournament using the codebook. The research assistants had no knowledge of the research question or the hypotheses. Disagreements among the coders were resolved by majority vote: an idea had to be assigned to a category by at least two research assistants for that to be considered a valid categorization.

The coders had little difficulty assigning ideas to codes. There were only 6 ideas (out of a total of 1123) for which there was no consensus between the research assistants. Out of all potential decisions, the research assistants agreed over 97% of the time on average.

Coders were also asked to identify ideas that should potentially be disqualified, either because they might be written by bots, or they were ideas that were too short to be meaningful (e.g. "sell shoes"). Ideas that received two or more disqualification votes were removed from the data. 1100 ideas received 0, 53 ideas received 1, 40 ideas received 2, and 70 ideas received 3 disqualify votes. Hence, in total, the research assistants voted to remove 110 ideas from the data.

After removing the ideas that received two or more disqualify votes by the research assistants, we went through the whole data set and read the ideas once again. We paid close attention to the ideas that received a single disqualify vote. After reviewing the ideas, we disqualified an additional 19 ideas, most of which were among the 53 ideas that received a single vote to be disqualified. As a result, 129 ideas were removed from the data set in total.

5.2.1 Study 2 Dependent Variables

We compute the distance between pairs of ideas in the same way as in Study 1. As before, we present analyses of the mean distance computed across all pairs of ideas for a participant and mean distance computed across the three most distant ideas. These measures are computed across all of the ideas submitted in the first stage of Study 2, i.e., prior to the participant choosing which ideas should be submitted for evaluation.

In the second stage of Study 2, participants were asked to select up to three of their ideas from the first stage for submission and evaluation. The mean distance between the pairs of ideas in this selected subset constitutes our third dependent variable in Study 2.

5.3 Study 2 Results

[TABLE 6 GOES HERE]

Table 6 summarizes the descriptive statistic for the three different measures of the variety of ideas. Note that the mean distance across the ideas selected for evaluation is only minimally different from the mean distance computed across all pairs of ideas, suggesting that participants did not prioritize variety in choosing what to submit to the evaluator.

The distribution of our measures of the variety of ideas by the experimental conditions can be found in Figure 5. This figure suggests that the greatest variety of ideas are found in the Anonymous Crowd condition, while there is little observable difference in the variety of ideas generated in the Expert and Known Crowd conditions.

[FIGURE 5 GOES HERE]

Table 7 presents results of OLS regressions with three dummy variables, the constant is the *Anonymous Crowd* condition while *Known Crowd* and *Expert* conditions are the other regressors.

[TABLE 7 GOES HERE]

Table 7 demonstrates that the anonymous crowd condition has higher point estimates for all three dependent variables. This reinforces the conclusion from Figure 5 that the anonymous crowd treatment results in a greater variety of ideas compared to the known crowd and expert treatments.

In other words, the subjects in the anonymous crowd condition generated and selected ideas that are more different than each other on average. As in Study 1, these effects are substantively meaningful with the expert condition reducing the variety of ideas by almost half a standard deviation relative to the anonymous crowd condition.

Furthermore, known crowd and expert conditions are almost identical in terms of point estimates and standard deviations for every regression. Indeed, F-tests failed to detect any statistically significant difference between the known crowd and expert conditions.For column 1 of Table 7, the F-statistic is 0.002; for column 2, 0.621; and for column 3, 0.189. None of these statistics are near significant at conventional levels.

Referring back to Table 1, the OLS regressions and the F-tests indicate that the difference between organizational and crowd selection results from *Identity Information* rather than *Number of Evaluators*. Although the known crowd and expert treatments differ on *Number of Evaluators*, their effects on participants are indistinguishable. The anonymous crowd treatment, on the other hand, is distinct from the other treatments. We conclude that *Identity Information* is the factor that causes the difference between organizational and crowd selection.

5.3.1 Idea Generation vs. Selection

One of the goals of study 2 was to understand whether the difference in idea variation occurs during idea generation or idea selection. The results in the first column of Table 7 indicate that the effects of the selection mechanism are apparent already at the idea generation stage. Recall that the mean distance measure in the first column of Table 7 was computed across all pairwise combinations of ideas, including those not selected for evaluation by the participant. Therefore, any difference in this measure cannot be explained by the participant strategically choosing ideas for the evaluator.

We can gain further insight into whether participants strategically emphasize variety when submitting ideas for evaluation. A simple way to do so is to compute the ratio of the mean distance across the ideas selected for evaluation to the mean distance among the three most distant ideas generated. The ratio indicates how much subjects diverge from the maximum possible distance when selecting their ideas; if they are emphasizing variety at the submission stage, this ratio should be close to one. However, a comparison of these ratios across the three conditions revealed no significant differences.

We conclude that the influence of the selection mechanisms on the variety of ideas occurs during the initial idea generation process. The treatments do not influence people during the explicit idea selection process.

6 Discussion

The results of Studies 1 and 2 indicate that when creators face a large number of evaluators about whom they lack identity information, they generate a greater variety of ideas than when they face single or multiple evaluators whom they know are experts in a specific domain. We see this as evidence of what we have called taste targeting: people's tendency to restrict their idea search according to their knowledge of the evaluator, and the expected selection criteria. A lack of identity information and a large number of evaluators (the anonymous crowd condition) makes it harder for people to develop a clear conception of evaluators' taste. As a result, the ideas they create exhibit less variety.

An important scope condition for these results is that they are observed within the context of an innovation tournament. This means that idea creators were explicitly motivated, through the promise of a bonus payment, to have their idea approved by an evaluator. The results suggest that under these conditions, the presence of identity information induces individuals to approach the creation of ideas as a coordination problem, or a matter of how best to coordinate with the evaluator (Weber and Camerer 2003). In the absence of such information, they approach idea creation as a search problem and cast a wider net, leading to a greater variety of ideas and an improved basis for exploratory learning.

Innovation tournaments are common in organizations, whether they are explicitly designed (e.g., with cash awards for successful ideas), or whether the rewards are implicit in terms of recognition, superior performance evaluations, and promotions. Yet not all organizations use such tournaments, and not all organizational creativity occurs within them. Employees may generate new ideas without the hope or promise of explicit rewards. Similarly, in many cases creativity may be affected by fear of sanctions rather than positive rewards. Such conditions are beyond the scope of our experiments, but future research should consider whether selection mechanisms (or beliefs about selection mechanisms) have similar effects under these conditions.

One generalizability concern with our studies is that idea creators in organizations have domain expertise, while the participants in our experiments do not (or have it only by chance). This may be particularly relevant to the operation of the anonymous crowd condition, where the idea creators in our experiments lack identifying information on the evaluators. Employees with domain expertise may, even if presented with an anonymous crowd selection technique, substitute assumptions about the domain expertise of the evaluators to match that of their employer. The extent to which such biases exist and can be overcome through the design of the selection mechanism is an important topic for future research.

7 Conclusion

The growing prevalence of crowd-based innovation methods has led to a robust and intriguing literature on the conditions under which such methods are effective for sourcing ideas and selecting among them. The growth of crowd-based innovation methodologies presents firms with a choice about how to design their innovation processes, a choice that corresponds to classic questions about vertical integration. Thus firms can choose to rely on internal mechanisms for both the variation and selection stages, rely on external mechanisms for both, or mix and match. Our study contributes to this literature by suggesting that crowd selection methods have important "upstream" consequences for the variation process. Firms that choose to generate variation internally may see a greater variety of ideas to the extent that they rely on external crowd selection rather than internal selection methods.

Our studies also shed light on the forces that might affect the balance between exploration and exploitation within firms. These forces arise because variation and selection processes are more likely to be reciprocally interdependent when both are performed within a firm. This reciprocal interdependence is particularly likely to be true when the variation and selection functions map on to different roles within the organization (Mollick 2012; Berg 2016), or in other words when the people who are expected to create the ideas are different from the people who choose which ideas are worth pursuing. Berg (2016) suggests that managers are worse at forecasting the success of others' ideas than creators. One implication of the poorer forecasting abilities of managers is that managerial selection will drive out potentially successful ideas. Our results suggest that managerial selection also has implications for innovation, but in terms of the variety of ideas that are generated and that form the basis for the organization's forecasting.

It is important to note, in this respect, that while our experimental manipulations of identity information and known evaluators maps on to a distinction between internal and external selection, firms have a fair degree of latitude in varying these factors when designing selection processes. In other words, the variety of ideas generated within a firm may be more a function of *how* internal selection methods are implemented than whether a firm uses managerial or crowd selection *per se*.

References

Amabile, T. M., R. Conti, H. Coon, J. Lazenby, and M. Herron (1996). Assessing the work environment for creativity. *Academy of Management Journal 39*(5), 1154–1184.

Amabile, T. M. and M. G. Pratt (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning.

Anderson, N., K. Potočnik, and J. Zhou (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management 40*(5), 1297–1333.

Barnett, W. P. and G. R. Carroll (1995). Modeling Internal Organizational Change. *Annual Review of Sociology 21*(1), 217–236.

Benner, M. J. and M. L. Tushman (2002). Process Management and Technological Innovation: A Longitudinal Study of the Photography and Paint Industries. *Administrative Science Quarterly 47*(4), 676–706.

Berg, J. M. (2016). Balancing on the Creative Highwire: Forecasting the Success of Novel Ideas in Organizations. *Administrative Science Quarterly 61*(3), 433–468.

Burt, R. S. (2004). Structural Holes and Good Ideas. *The American Journal of Sociology 110*(2), 349–399.

Collis, D. (2016). Lean strategy. *Harvard Business Review* (March), 1–14.

Ederer, F. and G. Manso (2013). Is Pay for Performance Detrimental to Innovation? *Management Science 59*(7), 1496–1513.

Felin, T., K. R. Lakhani, and M. L. Tushman (2017). Firms, crowds, and innovation. *Strategic Organization 15*(2), 119–140.

Girotra, K., C. Terwiesch, and K. T. Ulrich (2010). Idea Generation and the Quality of the Best Idea. *Management Science 56*(4), 591–605.

Knudsen, T. and D. A. Levinthal (2007). Two Faces of Search: Alternative Generation and Alternative Evaluation. *Organization Science 18*(1), 39–54.

Kolko, J. (2015). Design Thinking Comes of Age. *Harvard Business Review* (September), 1–9.

Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science 43*(7), 934–950.

Levinthal, D. A. (2007). Bringing selection back into our evolutionary theories of innovation. In F. Malerba and S. Brusoni (Eds.), *Perspectives on the Economics of Innovation* (First ed.)., Chapter 9, pp. 293–307. Cambridge, UK: Cambridge University Press.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science 2*(1), 71–87.

Martin, R. L. (2011). The innovation catalysts. *Harvard Business Review 89*(6), 1–11.

McGrath, R. G. (2001). Exploratory Learning , Innovative Capacity and Managerial Oversight. *The Academy of Management Journal 44*(1), 118–131.

Mollick, E. (2012). People and process, suits and innovators: The role of individuals in firm performance. *Strategic Management Journal 33*(9), 1001–1015.

Mollick, E. and R. Nanda (2016). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science 62*(6), 1533–1553.

O'Reilly, C. A. and M. L. Tushman (2013). Organizational Ambidexterity: Past, Present and Future. *The Academy of Management Perspectives 27*(4), 324–338.

Rigby, D. K., J. Sutherland, and H. Takeuchi (2016). Embracing Agile. *Harvard Business Review* (May), 1–23.

Simonton, D. K. (2011). Creativity and Discovery as Blind Variation: Campbell's (1960) BVSR Model After the Half-Century Mark. *Review of General Psychology 15*(2), 158–174.

Singh, J. and L. Fleming (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science 56*(1), 41–56.

Sorensen, J. B. (2002). The Strength of Corporate Culture and the Reliability of Firm Performance. *Administrative Science Quarterly 47*(1), 70–91.

Sorensen, J. B. and T. E. Stuart (2000). Aging, Obsolescence, and Organizational Innovation. *Administrative Science Quarterly 45*(1), 81–112.

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

Weber, R. a. and C. F. Camerer (2003). Cultural Conflict and Merger Failure: An Experimental Approach. *Management Science 49*(4), 400–415.

Weinberg, J. D., J. Freese, and D. McElhattan (2014). Comparing data characteristics and results of an online factorial survey between a population-based and a crowdsource-recruited sample. *Sociological Science 1*, 292–310.

Yuan, F. and J. Zhou (2008). Differential Effects of Expected External Evaluation on Different Parts of the Creative Idea Production Process and on Final Product Creativity. *Creativity Research Journal 20*(4), 391–403.

Zhou, J. (2003). When the presence of creative coworkers is related to creativity: Role of supervisor close monitoring, developmental feedback, and creative personality. *Journal of Applied Psychology 88*(3), 413–422.

Figures and Tables

Figure 1: Taste Targeting

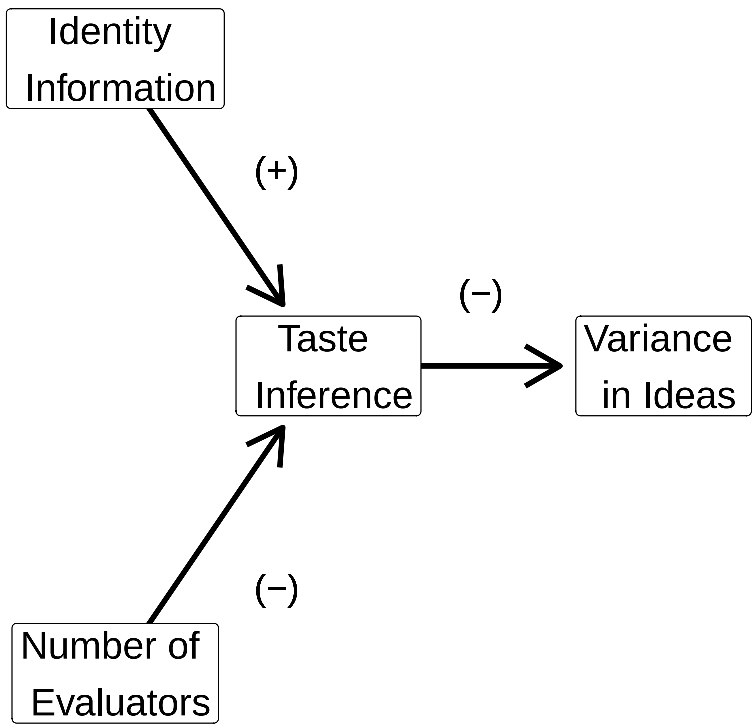


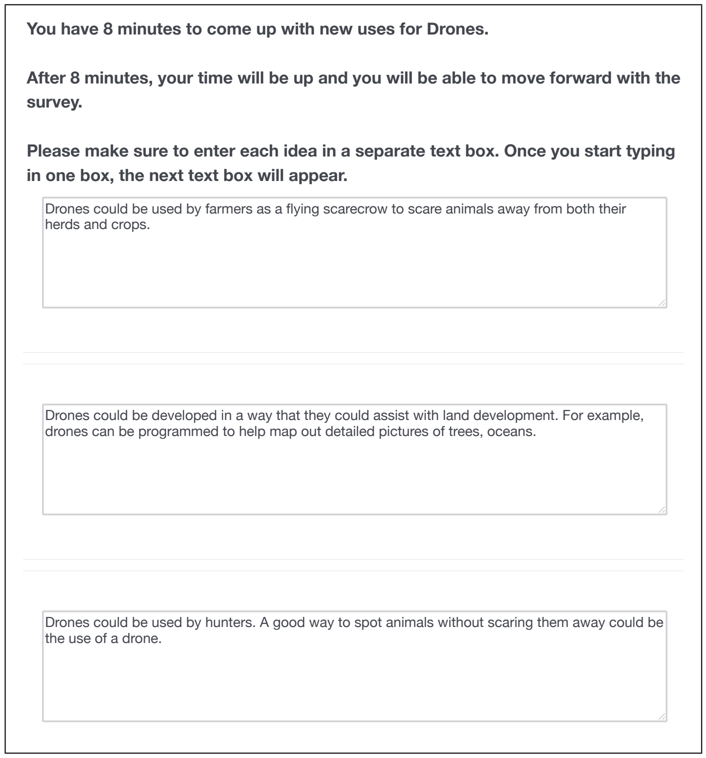
Table 1: Evaluator Characteristics

		Number of Evaluators	
		Few	**Many**
Identity Information	**Yes**	A	B
	No		C

Figure 2: The Expert Treatment in Study 1

Kevin Ferguson is an industry veteran in product marketing. After receiving his MBA from Stanford Graduate School of Business, he began his career as a marketing manager for Microsoft where he focused on how to expand Microsoft Office to new customer groups. Later, he joined Amazon as a marketing director. Currently, Kevin advises start-ups on how to expand their customer base by employing their products in different areas.

Figure 3: Idea Entry Page



You have 8 minutes to come up with new uses for Drones.

After 8 minutes, your time will be up and you will be able to move forward with the survey.

Please make sure to enter each idea in a separate text box. Once you start typing in one box, the next text box will appear.

Drones could be used by farmers as a flying scarecrow to scare animals away from both their herds and crops.

Drones could be developed in a way that they could assist with land development. For example, drones can be programmed to help map out detailed pictures of trees, oceans.

Drones could be used by hunters. A good way to spot animals without scaring them away could be the use of a drone.

Table 2: Descriptive Statistics - Study 1

Statistic	Anonymous Crowd	Expert
Number of Participants	71	73
Number of Ideas	452	372
Mean N of Ideas per Person	6.36	5.09
Std. Dev.	3.22	2.59
Mean N of Characters per Idea	119.00	135.00
Std. Dev.	101.00	107.00

Table 3: Measures of the Variety of Ideas in Study 1

Mean Distance Across	Mean	Median	St. Dev.	Min	Max
All Pairs of Ideas	3.62	3.70	0.67	0.00	5.00
3 Most Distant Pairs of Ideas	4.02	4.00	0.80	0.00	6.00

Figure 4: Variety of Ideas by Experimental Condition, Study 1

Note: The plots have different Y axes

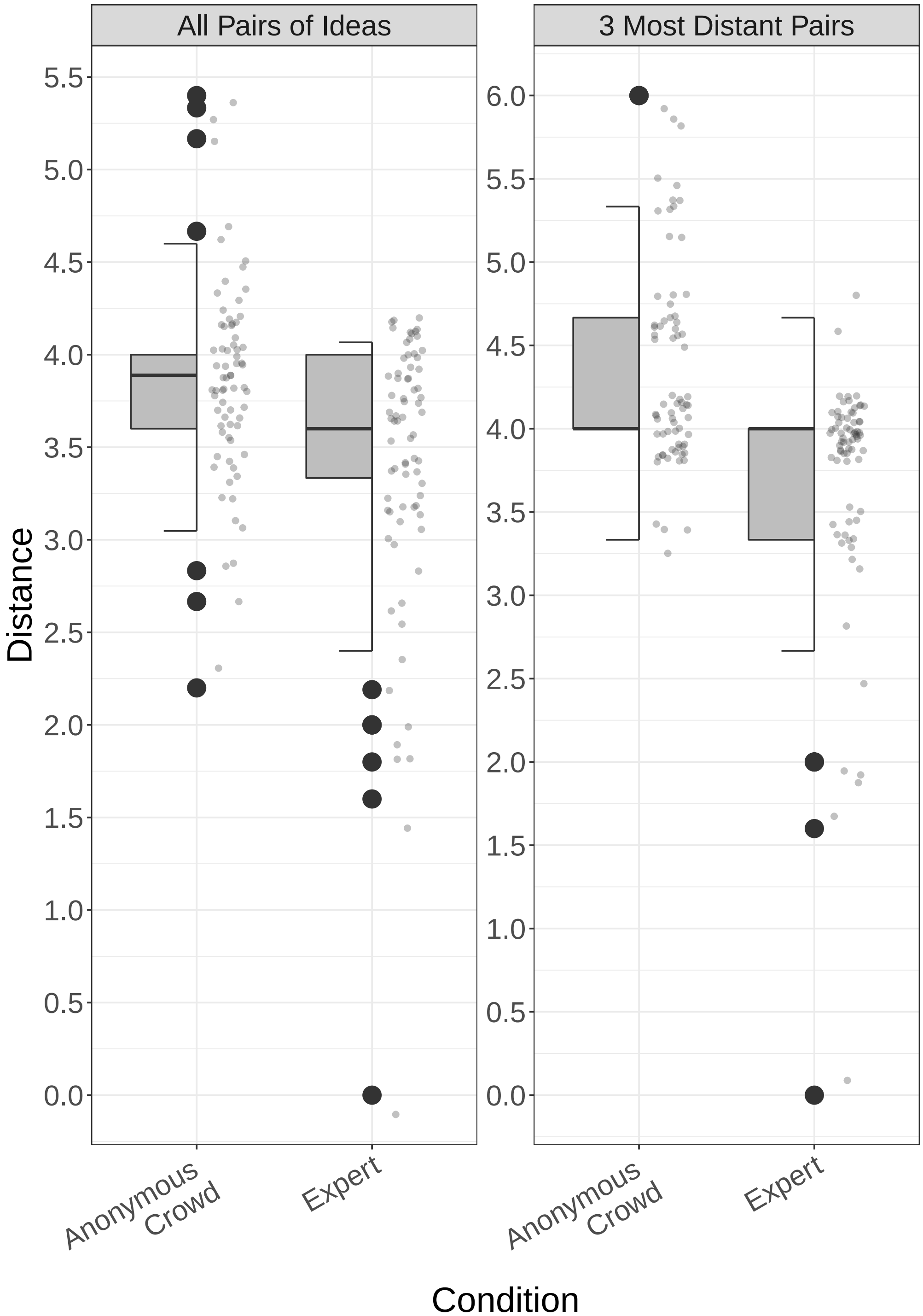


Table 4: OLS Regressions of Variation in Ideas on Selection Method, Study 1

	Mean Distance Across:	
	All Pairs of Ideas	3 Most Distant Pairs
	(1)	(2)
Anonymous Crowd	3.854*	4.385*
	(0.075)	(0.085)
Expert	−0.463*	−0.714*
	(0.105)	(0.120)
Observations	144	144
R^2	0.120	0.199
Adjusted R^2	0.114	0.194
Residual Std Error (df = 142)	0.630	0.720
F Statistic (df = 1; 142)	19.402*	35.383*

Note: *p<0.01

Table 5: Descriptive Statistics - Study 2

Statistic	Anonymous Crowd	Expert	Known Crowd
Number of Participants	62	56	75
Number of Ideas	393	302	415
Mean N of Ideas per Person	6.34	5.39	5.53
Std. Dev.	2.59	2.61	3.00
Mean N of Characters per Idea	69.70	70.10	58.50
Std. Dev.	64.10	71.10	51.40

Table 6: Measures of the Variety of Ideas in Study 2

Mean Distance Across	Mean	Median	St. Dev.	Min	Max
All Pairs of Ideas	3.08	3.20	0.91	0.50	5.00
3 Most Distant Pairs of Ideas	3.70	4.00	1.11	0.50	6.66
Ideas Selected for Evaluation	3.11	3.33	1.03	0.50	5.33

Figure 5: Variety of Ideas by Experimental Condition, Study 2

Note: The plots have different Y axes

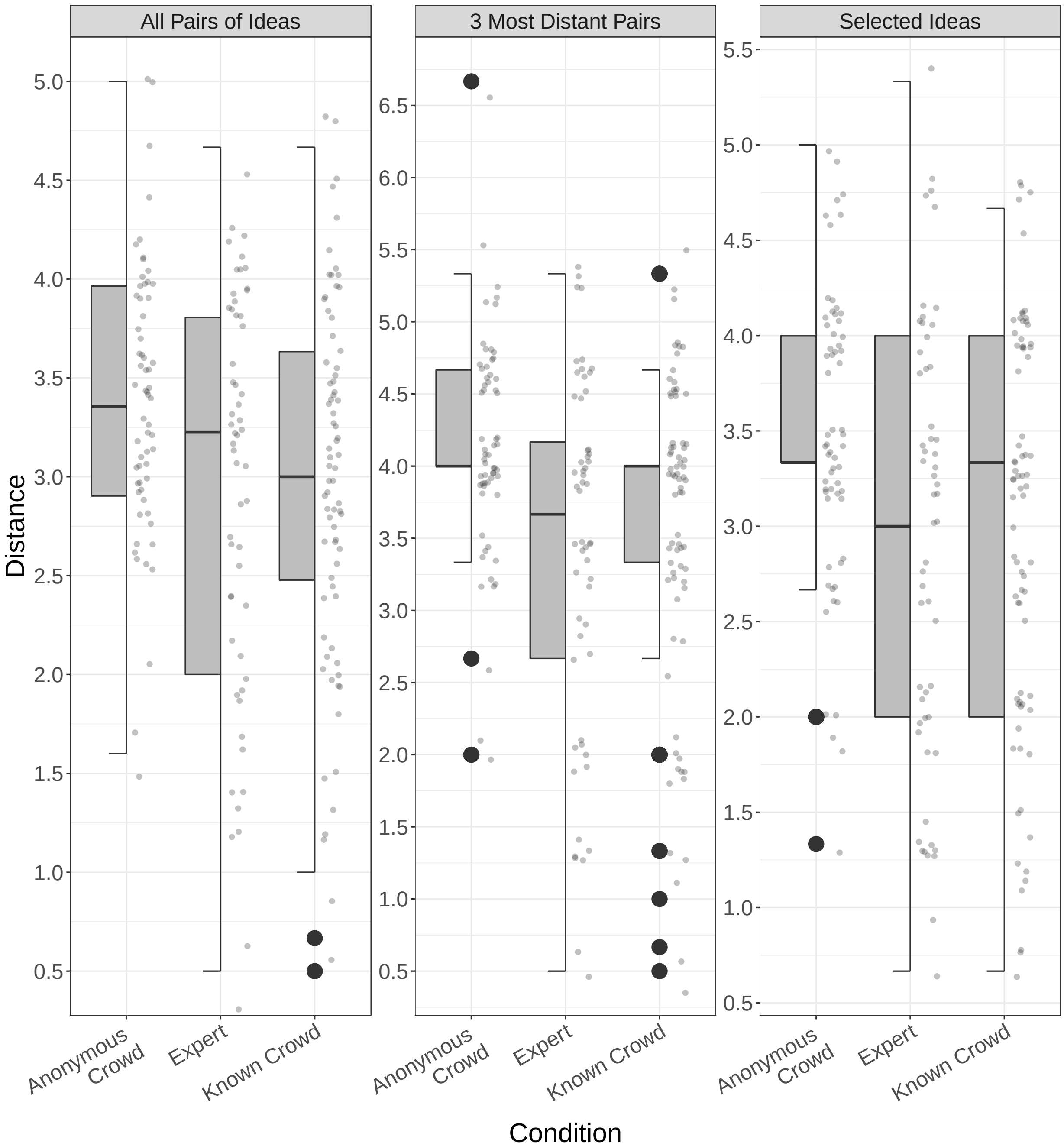


Table 7: OLS Regressions of Variation in Ideas on Selection Method, Study 2

	Average Distance Across:		
	All Pairs of Ideas	3 Most Distant Pairs	Selected Ideas
	(1)	(2)	(3)
Anonymous Crowd	3.373*	4.161*	3.482*
	(0.114)	(0.136)	(0.128)
Expert	−0.431*	−0.757*	−0.589*
	(0.166)	(0.198)	(0.185)
Known Crowd	−0.412*	−0.601*	−0.508*
	(0.154)	(0.184)	(0.173)
Observations	193	193	193
R^2	0.046	0.082	0.062
Adjusted R^2	0.036	0.072	0.052
Residual Std. Error (df = 190)	0.898	1.075	1.005
F Statistic (df = 2; 190)	4.618†	8.454*	6.239*

Note: $^{\dagger}p<0.05$; $^{*}p<0.01$