INSEAD

The Business School for the World®

Faculty & Research Working Paper

Lone Inventors as Sources of Breakthroughs: Myth or Reality?

Lee FLEMING
Jasjit SINGH
2008/69/ST

Lone Inventors as Sources of Breakthroughs: Myth or Reality?

By

Lee Fleming*

And

Jasjit Singh**

October 15, 2008

We would like to thank the Harvard Business School Division of Research and INSEAD for supporting this work; Henrich Greve, Santiago Mingo, and Scott Stern for their comments; and participants in presentations at Stanford, Harvard, the University of Toronto, and the Academy of Management and INFORMS annual conferences for their feedback. Christoph Loch and two anonymous reviewers provided exceptional help in developing the work.

* Professor of Business Administration at Harvard Business School 485 Morgan Hall Harvard Business School Boston, MA 02163 Ph: (617) 495-6613 Email: lfleming@hbs.edu

** Assistant Professor of Strategy at INSEAD, 1 Ayer Rajah Avenue, Singapore 138676 Ph: +65 6799 5341 E-Mail: jasjit.singh@insead.edu

Abstract

How does collaboration influence creativity and, in particular, the invention of breakthroughs? Recent research has attempted to resolve this question by considering the variance of creative outcomes, implicitly assuming that greater probability of breakthroughs comes at the cost of greater probability of particularly poor outcomes. However, through an examination of the overall distribution of outcomes in the context of patented inventions, we demonstrate that collaboration can have opposite effects at the two tails of the distribution: it reduces the probability of very poor outcomes while simultaneously increasing the probability of extremely successful outcomes. We find that these effects are at least partially mediated by the technical diversity of team members and by the size of team members' external collaboration networks. We also find that large teams exhibit greater extent and breadth of technological search than small teams or lone inventors, and that such search behavior is associated with greater impact.

Keywords: Creativity, collaboration, invention, teams, creativity, quantile, variance

INTRODUCTION

Are inventors who work alone more creative than those who collaborate—and more likely to invent breakthroughs? Investigations of creativity have described many advantages of collaboration (Sutton and Hargadon 1996; McFadyen and Cannella 2004; Wuchty, Jones, and Uzzi 2007). Yet a compendium on social creativity acknowledged that the research basis for the benefits of collaboration remains "somewhat weak" (Paulus and Nijstad 2003: 4); related research illuminates many problems of working with others (Diehl and Stroebe 1987; Mullen, Johnson, and Salas 1991; Runco 1995; Paulus and Brown 2003). The more focused question of who is more likely to invent a breakthrough also remains unsettled. Inventors who work alone (often called "lone inventors" [1]) are often cited as the sources of discontinuous breakthroughs (Schumpeter 1934; Tushman and Anderson 1986; Mokyr 1990: 295; Hughes 2004: 53), yet others have argued for the iterative and social continuity of creative search (Gilfillan 1935; Schumpeter 1942; Basalla 1988). Accounts of lone heroic breakthroughs remain common in journalism and literature (Morison 1966; Schwartz 2002; Evans 2005); Nobel-prizewinning author John Steinbeck (1952: 130-131) wrote:

> Our species is the only creative species, and it has only one creative instrument, the individual mind and spirit of a man. Nothing was ever created by two men. There are no good collaborations, whether in music, in art, in poetry, in mathematics, in philosophy. Once the miracle of creation has taken place, the group can build and extend it, but the group never invents anything. The precociousness lies in the lonely mind of a man.

The goal of this paper is to enlarge the solution space for these debates, particularly with regards to the sources of breakthroughs and the processes by which they are conceived and developed. We argue (and provide supporting evidence from patent data) that research on collaboration and creativity should theorize about and examine the complete distribution of creative outcomes. Following most work in statistical theory and estimation, almost all research on creativity has considered the influence of explanatory variables on the average or mean outcome. Recent work, however, has begun to empirically model the second moment or variance of distributions of creative outcomes (Dalhin, Taylor, and Fichman

[1] It is important to differentiate between inventors who do not collaborate and inventors that do not work for a firm. We control for the latter characteristic and focus on the influence of collaboration.

2004; Taylor and Greve 2006; Fleming 2007; Girotra, Terwiesch, and Ulrich 2007). This work has been theoretically motivated by an interest in particularly creative outcomes ("breakthroughs"). Greater variance is preferred in the quest for breakthroughs (March 1991), because it increases the mass in the tails and the increased mass in the upper tail implies a greater number of breakthrough outliers.

Unfortunately, the results of this research on the mean and variance of creativity are inconsistent. Dahlin, Taylor, and Fichman (2004) demonstrate that independent inventors (those who do not work for a firm) are overrepresented in the tails of creative distributions; because independent inventors are more likely to work alone, this implies that collaborative inventors will be underrepresented in the tails. Consistent with this evidence, Fleming (2007) uses mean-variance decomposition models (King 1989) to estimate a lower average and greater dispersion of creative outcomes by individuals who work alone. In contrast to these results, Taylor and Greve (2006) demonstrate higher variance in deviation from a normalized mean measure. Girotra, Terwiesch, and Ulrich (2007) adopt an experimental approach and likewise demonstrate higher variance in outcomes for creative collaborations. We develop this line of research on creativity and the higher moments and reconcile the conflicting results by demonstrating differential effects of collaboration on the upper and lower tails of creative distributions.

We argue that an excessive focus on variance of outcomes, in addition to being inconclusive, can also mislead researchers in understanding how collaboration affects the overall distribution of outcomes from the creative process. The hypothetical cumulative distribution function (CDF) of Figure (1) illustrates several ways in which collaboration could affect the distribution of outcomes, even assuming that it is beneficial on an average (though some would dispute the assumption; see Paulus and Nijstad 2003). The debate on individuals versus collaborative teams as a source of breakthroughs is often framed as whether the observed variance of outcomes is more in line with scenario (1) or scenario (2) in the figure. If individual inventors are associated with greater variance of outcomes, scenario (1) is implicitly assumed and collaboration is therefore judged as being less desirable if the goal is to achieve breakthrough

innovations. On the other hand, if individual inventors are associated with lower variance, then scenario (2) is implicitly assumed and collaboration is judged as being desirable for breakthrough outcomes. Both arguments assume symmetry; it is not considered possible to simultaneously increase the probability of particularly good outcomes and decrease the probability of particularly bad outcomes.

We argue that the above reasoning just presented, though recently common, is incomplete and potentially misleading. In particular, as scenarios (3) and (4) in Figure (1) illustrate, achieving greater variance is neither necessary nor sufficient for ensuring greater likelihood of breakthroughs. While collaboration is associated with lower variance in scenario (3) and greater variance in scenario (4), it is better at achieving breakthroughs in both scenarios. Further, the effects at the two tails need not involve a trade-off; collaboration can potentially increase the likelihood of breakthroughs while simultaneously reducing the probability of particularly bad outcomes.[2]

Building on a stylized evolutionary model of creativity, we argue that collaborative inventors are less likely to invent and develop poor outcomes *and* more likely to invent and develop highly skewed outliers. Supporting an argument that collaboration improves the sorting and identification of promising new ideas, we find that working with others trims the lower tail of the distribution of creative outcomes. Supporting an argument that collaboration also enables more creative novelty, we find that working with others also shifts the mass of creative outcomes towards extremely successful outliers. These beneficial effects on the distribution of outcomes reflect more than an upward mean shift; the effect is found to depend significantly on the quantile of the outcome distribution (Koenker and Bassett 1978). We then use mediation models to further probe the mechanisms and processes that produce these results. The

[2] The illustrative example should not be taken literally. The scenarios in Figure 1 have been generated by assuming a normal distribution where collaboration increases the mean but is allowed to increase or decrease the variance to different extents. Conceptually, the distribution with larger variance *will* in the limit have a greater probability of the very extreme outcomes at both ends. However, the probability mass where this happens might be too trivial to be of economic significance (as demonstrated by Fleming, 2007). In addition, real-world outcomes need not obey a normal—or symmetric—distribution.

diversity of a collaborative team and the size of its external networks partially mediate the benefits of collaboration, consistent with the argument that diversity and external networks enable both greater recombinant novelty and more rigorous assessment of that novelty. Furthermore, teams invent technologies of greater breadth and scope, characteristics which also partially mediate the advantages of collaboration.

EVOLUTIONARY THEORIES OF CREATIVITY

Following many researchers (Campbell 1960; Romer 1993; Weitzman 1996; Simonton 1999), we view creativity as an evolutionary search process across a combinatorial space. In the first phase of evolutionary search, typically called the "variation" phase, people generate new ideas through combinatorial thought trials. In the second or "selection" phase, they evaluate these ideas in order to reject poor outcomes and identify the most promising novelties. The processes within these two phases can be purely psychological—that is, they can all occur within a single person—or they can iterate between psychological and social-psychological processes. The purely psychological case would be an extreme example of a lone inventor who has no interaction with collaborators or feedback of any kind. This archetypal example is probably quite rare in today's interconnected world. In the latter case, at the other extreme of social-psychological processes, individuals work together closely in both the generation and evaluation of ideas (though, following Steinbeck, we believe that each generative insight occurs within a single mind, after which the insight may be shared, iterated on, and further recombined by collaborators). In this model, most creative search in its early stages iterates between generation and selection, during which time the inventors select very few of their new combinations for development.

In the last or "retention" phase, members of a larger creativity community evaluate the selected ideas and go on to adopt a very few of them in their own creative searches. This phase is mainly social. Indeed, except in the very rare cases when a purely objective measure of the quality of an idea or invention can be used, it is completely social. While objective measures may be possible in a univariate analysis of a

particular technology characteristic (such as transistor density or miles per gallon), it is difficult to assess an intrinsic and completely asocial value for most technologies and even more difficult to make comparisons across technologies. Even "expert" assessment is still social, as the inventor(s) must necessarily communicate the idea to the experts. Many creativity researchers have developed these arguments; Csikszentmihalyi (1999), for example, argues: "To be creative, a variation must somehow be endorsed by the field…Creativity involves social judgment." (See also Simonton 1999 and Paulus and Nijstad 2003.) Creative individuals can incorporate their own prior work, but their influence will be limited unless others pick up and build on their ideas. Following this evolutionary model, we define the ultimate success of a new idea as its impact on the future invention efforts of others (our results are robust to the inclusion of the inventor's own further efforts).

EMPIRICAL EXPECTATIONS

The idea that novelty is a new combination is at least as old as Adam Smith (1766, reprinted 1982). Given a thorough enough historical search, novel technologies can almost always be traced to combinations of prior technologies (Basalla 1988) and science, music, language, art, design, manufacturing, and many other forms of creative endeavor have been described similarly (Gilfillan 1935; Romer 1993; Weitzman 1999). Contriving a new combination is just the initial variation stage, however, and is usually followed quickly—in the selection phase—by consideration of the new combination's worth. Independent of the idea's source, we propose that collaboration improves the effectiveness of the selection phase because collaborative selection will be more rigorous.

Returning to a stylized evolutionary model, we assume that individual inventors create and then immediately test their ideas and new combinations within their own minds (Campbell 1960). Most new ideas are quickly rejected; only a few are retained as the basis for continued search. The quality of even the retained ideas is still suspect, however, because individuals, whether experts (Simonton 1985) or non-experts (Runco and Smith 1991), are notoriously bad evaluators of their own ideas. Teams have an

inherent advantage in the identification of the best ideas. A collaborative team will consider the invention from a greater variety of viewpoints and potential applications; such broader consideration is more likely to uncover problems. Given the typically greater diversity of experience on a collaborative team, some member is more likely to recall having seen a problem with a similar invention and argue to abandon or modify the approach. In short, collaborative creativity will subject individually conceived ideas to a more rigorous selection process, so that fewer poor ideas are pursued.

Anecdotal accounts by prolific and successful lone inventors support the argument; such inventors readily admit and joke about their inability to predict which of their inventions would prove to be breakthroughs (Schwartz 2004: 144). They often report a division of labor between those who generate and those who criticize: "You wanted Charlie in the conversation, because he would tell you when you were full of it" (Kenney 2006). The inventor of the aluminum tennis racket, Styrofoam egg cartons, and plastic milk bottles reported that "the problem with the loner is that if you don't sift, you are liable to spend much time going down dead ends" (Brefka 2006). Referring to the inventor of a promising automated language-translator, a Carnegie-Mellon professor reports that "Eli is a mad genius… Both of those words apply. Some of his ideas are totally bogus. And some of his ideas are brilliant. Eli himself can't always tell the two apart" (Ratliff 2006). The dual inventors of the Hewlett Packard thermal ink-jet were a prolific empirical tinkerer who generated prototypes and a very methodical engineer who explained, documented, and criticized (Fleming 2002).

Because collaborative teams will provide additional and (to varying degrees) uncorrelated filters, they will uncover more potential problems and hence develop fewer dead ends. The individual inventor, lacking the advantage of collaborative sorting, will develop more poor ideas, with the result that a smaller proportion of her developed ideas will be used by others. Hence, we would expect collaboration to "trim" the lower tail of the distribution of creative outcomes.

H1: The distribution of ideas created by collaborations will be shifted away from lower-impact outcomes, relative to the distribution of ideas created by individuals.

In addition to trimming the undesirable tail, collaboration will also fatten the desirable end of the distribution. Repeating the oft-cited advantages of diversity (Gilfillan 1935; Basalla 1988; Weitzman 1999), we argue that collaboration increases the potential combinatorial opportunity for creating novelty. Each inventor brings a different set of past experiences and knowledge of potential technologies to the search and thus increases the potential number of new combinations that can be generated. New combinations are more uncertain and more variable in their impact (Fleming 2001). In other words, they should increase the likelihood of both good and bad outcomes. Following the arguments for Hypothesis 1, however, collaboration should also provide a more rigorous selection process, so the worse outcomes should still be less likely to be developed.

Collaborative teams also generate more points in the upper tail because they can cycle through a greater number of iterations. Particularly if they work well and productively together, they can quickly generate and assess many potential options. On the assumption that teams invest more total effort in aggregate, we can expect them to iterate more quickly in the generation and selection phases of creative search, thus generating more possible breakthroughs and avoiding more poor outcomes.

For both these reasons, greater diversity that enables greater novelty and a greater volume of iterations, collaborative teams can generate more potential novelty at the breakthrough end of the distribution.

H2: The distribution of ideas created by collaborations will be shifted towards higher-impact outcomes, relative to the distribution of ideas created by individuals.

Note that these predictions do not depend on the influence of collaboration on the average outcome. While we believe that collaboration should also influence the mean positively (consistent with McFadden and Canella 2004), we argue that our predictions involve more than a symmetric and upward shift of the

mean (as could be realized by adding the same offset to every point in a distribution). We will establish this empirically by demonstrating effects of different sizes on the two tails, in both logit models of extreme outcomes and quantile regressions.

The arguments for the first two hypotheses depend heavily on the role of diversity. Diversity of team backgrounds helps generate greater recombinant novelty and potential applications, thus fattening the upper tail. At the same time, diversity helps identify a poor outcome before it is fully developed, thus trimming the lower tail as well. These arguments imply that diversity mediates the value of collaboration; if they are correct, then the observed benefits of collaboration decrease when the extent of diversity is explicitly accounted for in the model (Baron and Kenney 1986).

H3: The effect of collaboration on the distribution of innovative outcomes will be mediated by the diversity of backgrounds of the collaborators.

The arguments for the value of collaboration should also apply to indirect collaborators—people who work with one or more of the team members on another project but are not a part of the immediate effort. These extended or "supporting" team members should provide benefits similar to those provided by the immediate team members. Such colleagues act as additional sources of recombinant diversity and should therefore enhance the number of outcomes at the top end of the distribution. They also act as additional filters to trim the bottom end. For example, when an immediate team member describes a project to non-team colleagues, they can suggest overlooked possibilities and problems.

As with diversity, the extended network of collaborators should mediate the value of collaboration. Hence, when the size of collaborators' extended networks is explicitly accounted for in the model, there should be a significant decrease in (a) the negative correlation between collaboration and low impact and (b) the positive correlation between collaboration and high impact (Baron and Kenney 1986)

H4: The effect of collaboration on the distribution of innovative outcomes will be mediated by the size of the extended social network of the collaborating team.

DATA

We examine the link between collaboration and the distribution of creative outcomes using patent data. We follow the well-established tradition of using the extent to which a specific patent gets cited by future patents as a measure of its impact and, ultimately, its success.[3] Admittedly, citations are noisy as a measure of a patent's value, since they are used by a variety of parties for different reasons (Alcacer and Gittelman 2006). The number of citations a patent receives has nevertheless been shown to be correlated with several direct measures of patent value, including the consumer surplus generated (Trajtenberg 1990), expert evaluation of patent value (Albert et al. 1991), patent renewal rates (Harhoff et al. 1999), and contribution to a firm's market value (Hall et al. 2005).[4]

Since we are interested not only in the mean effect of collaboration but in the entire distribution of outcomes (and extremely rare breakthrough outliers in particular), we require large samples of comparable outcomes in order to say anything conclusive. Patent data are particularly suited for such an analysis, providing the time span and breadth necessary to identify breakthroughs and observe entire distributions of creative outcomes.[5] The sample we analyze in our regression analysis consists of patents filed for with the U.S. Patents and Trademarks Office (USPTO) during the 10-year period 1986-1995. We restricted ourselves to patents filed by U.S.-based inventors for comparability of their citation impact and to avoid much greater inaccuracies in the identification of foreign inventors. However, for constructing

---

[3] Admittedly, firms and individuals differ quite a lot in their reliance on patents, often relying on alternate means of protecting their intellectual property (Levin et al. 1987). This makes patent counts potentially misleading as measures of the overall extent of an innovation. Nevertheless, conditional on a specific innovation being patented, citations to that patent help capture its overall economic and technological success.

[4] This is also consistent with view of the Office of Technology Assessment and Forecast: "If a single document is cited in numerous patents, the technology revealed in that document is apparently involved in many developmental efforts. Thus, the number of times a patent document is cited may be a measure of its technological significance" (U.S. Patents and Trademarks Office (USPTO), Sixth Report, 1976, p. 167).

[5] In analyses not reported in the paper due to space limitations, we ensured the robustness of our findings to using three alternate approaches. The first was to examine the citation impact of collaborative versus individual researcher projects in the context of scientific publications rather than in the context of patents. The second was to examine the reuse of novel technology combinations over three-year periods of an inventor's career (following Fleming, Mingo, and Chen, 2007). The third used the latter dataset but estimated future citations to patents in the three-year window. Using different approaches to test the same hypotheses ensures the robustness of our findings to the idiosyncrasies of a particular dataset.

relevant measures such as prior experience of the inventors and future citations received by a patent, we do use data on all USPTO patents granted to inventors from anywhere in the world during the period 1975-2004.

Our dataset was constructed by merging data obtained directly from the USPTO and the National University of Singapore patent database as well as the National Bureau of Economic Research (NBER) database (Jaffe and Trajtenberg 2002: Chapter 13). Each patent record contains the patent number, the application and grant dates, the names and addresses of all inventors, a list of all technology classes and subclasses to which the patent pertains, and the assignee organization. Because inventors are not uniquely identified in the original data, we applied inventor-matching algorithms similar to those previously employed by Singh (2005; 2008), Trajtenberg (2006), and Fleming, Mingo, and Chen (2007) to create a reliable inventor-patent mapping.

A simple inspection of the data provides preliminary support for our arguments. As indicated in Table 1, collaboration appears to be associated with more citations, a benefit which appears to increase unambiguously with team size. The standard deviation of citation outcomes also increases with team size, though the trend is less clear if we employ the coefficient of variation as a scale-free measure of dispersion. The reported percentile statistics suggest that collaborative teams are disproportionately *more likely* to get breakthrough (i.e., high-citation) outcomes while simultaneously being *less likely* to get low-value (i.e., low-citation) outcomes. To understand how this is happening, it is useful to refer back to Figure 1, which illustrates four different scenarios that are possible even with the assumption that collaboration increases the mean of the distribution of possible outcomes. In our earlier discussion of Figure 1, we argued for the importance of examining how collaboration might affect the overall distribution; examining only variance can be misleading with regard to extreme outcomes. The raw data appear to be most consistent with scenario (4) in the figure.

Figure 2 illustrates the actual cumulative distribution function of the citation impact for different team sizes.[6] Notice that the graphs for different team sizes *never* intersect. Instead, the outcome distribution for larger teams stochastically dominates the distribution for smaller teams and for the lone inventor. Once more, this is consistent with a view that collaboration increases the probability of breakthroughs while decreasing the probability of particularly bad outcomes. Again, comparing the cumulative distribution function with the different possibilities shown in Figure 1, the distribution of outcomes for lone inventors versus collaborative teams in the actual data seems most consistent with scenario (4). The next section introduces regression models that enable us to estimate the magnitudes of these effects.

REGRESSION METHODOLOGY

As a baseline, we first analyze the effect of collaboration on the expected number of citations a random patent would receive. Since patent citations involve count data skewed to the right (and likely over-dispersed relative to a Poisson distribution), we employ a negative binomial model of the citations produced for each patent. Our key explanatory variable is *team_size*. Since patents from different years have different "windows of opportunity" to be cited, a direct comparison of patent citations across patents from different years would be inappropriate. Including year fixed effects in all regressions accounts for systematic cross-year differences arising from this "truncation bias." Similarly, technology dummies account for systematic differences in citation rate across patents from different technologies. To account for the possibility that the error terms are correlated for observations involving the same inventor, our estimation employs robust standard errors that are clustered on the identity of the first inventor.[7]

[6] To enable valid comparisons across different types of patents, the citation impact of each patent was first normalized relative to the average citation impact of the cohort of patents with the same application year and technology class. On account of the highly skewed nature of the citation impact, the variable was further transformed by taking its natural logarithm in order to make the chart more readable.

[7] One concern with this cross-sectional approach is that better inventors may have a greater propensity to engage in external collaboration. This could lead to an upward bias in estimated gains from collaboration. As discussed later,

The models include several control variables: *unassigned* (an indicator for whether the patent is unassigned, which typically indicates lack of corporate ownership), *university* (an indicator for whether the patent has a university assignee), *claims* (the scope of the patent as measured by its number of claims, assumed to correlate with the extent of the resources devoted to the project, and at least a partial control for the greater effort that teams can bring to an inventive effort), *inventor_age* (the average of the number of years since each team member applied for his or her first patent), and *team_patents* (the number of past patents from anyone in this team).[8]

The main focus of this paper, however, is to examine the effect of collaboration across the entire distribution of outcomes. We are particularly interested in the two extremes—inventions with little impact and inventions that can be considered breakthroughs. Hence, our next set of models is meant to examine the probability of particularly good and particularly bad outcomes. Specifically, logistic regression models are used to estimate the likelihood that a patent's impact lies in one of the tails of the distribution. Breakthrough innovations are measured using two dependent variables: an indicator for a patent ending up being in the top 1% in terms of frequency of citation in patents (among patents with the same application year and technology class), and a similar indicator for being among the top 5%. Analogously, two indicator variables are constructed to capture low-impact innovations: one for patents that receive no citations, and one for those that receive at most one citation. After examining the effect of collaboration at the mean as well as the two extremes of the outcome, we employ a quantile regression approach to systematically study how collaboration affects the entire distribution of outcomes.

---

we examined the robustness of our findings to employing panel data models with inventor fixed effects, at least partially addressing this concern.

[8] Collaboration might, of course, affect impact in part by changing some of these observed characteristics of the patent; e.g., by producing patents that have wider scope (more claims) or involve wider search (more backward citations). Including such characteristics as control variables therefore results in the benefits attributed to collaboration being more conservative. All findings reported in this paper, however, are qualitatively identical even if these controls are not used.

Having investigated the effect of collaboration across the entire distribution, we proceed to test the mediator hypotheses for why collaboration is beneficial. We measure the proposed mediation effects with *team_experience_diversity* (the number of different technology classes anyone in the team has patented in before) and *team_network_size* (the number of unique other collaborators or collaborators' collaborators for anyone in the team during the preceding five years).

RESULTS

A summary of definitions and key statistics for all variables appears in Table 2, and a correlation matrix among these variables appears in Table 3. Next, we report results from the regression analyses. Since many of our independent variables are highly skewed, we take natural logarithms of these variables before using them in the regressions.[9]

Collaboration and the impact of creative outcomes

Before examining the full distribution of outcomes, we formally examine the baseline expectation of a positive effect of collaboration on the average impact of creative outcomes. The positive estimate for *ln_team_size* in (1) of Table 4 supports this proposition, a finding that holds even after introducing relevant control variables in column (2). The magnitude of this effect is substantial: the estimates from column (2) imply a difference of 23% in the expected citation impact for lone inventors versus four-person teams. Estimates for the control variables are consistent with previous research: Patents with assignees, particularly those assigned to universities, correlate with greater impact; a greater number of claims is associated with more future citations; inventor age is negatively correlated with a patent's citation impact; and a larger cumulative team experience associates with greater impact.

[9] We first added one to those variables that can take a value of zero. The results are not substantively affected by the size of the offset or estimating with the raw variable.

Next, we investigate how collaboration affects the likelihood of getting particularly good or particularly bad outcomes. As columns (3) and (4) in Table 4 demonstrate, collaboration increases the probability of being a top 5% or top 1% patent in terms of citation impact, with larger teams having an even greater probability of such breakthroughs than smaller teams. The magnitude of the effect is substantial. For example, four-person teams are 64% and 65% more likely than lone inventors to invent a patent in the top 5% and top 1%, respectively. On the other hand, as columns (5) and (6) demonstrate, collaboration simultaneously *decreases* the probability of the patent ending up getting no citations or at most one citation. For example, four-person teams are 28% and 29% less likely than lone inventors to invent a patent with no citations or at most one citation, respectively. The evidence supports the argument that the effect of collaboration is asymmetric at the two extremes: collaboration *increases* the probability of particularly good creative outcomes and *decreases* the probability of particularly bad outcomes. Estimating a model of the second and/or first moments would miss this asymmetry.

Examining the role of team structure: Experience diversity and network size as mediators

We now dig deeper into mediation analysis examining why collaboration helps, and why larger teams are better. Following Baron and Kenny (1986), our mediation analysis takes three steps. The first step, omitted here in order to conserve space, established that *team_size* (the explanatory variable) significantly affects the mediators in the expected way. In other words, having larger teams is associated with greater *team_experience_diversity* and greater *team_network_size*. The second step, also not reported here, established that the above mediators significantly affect the three dependent variables - average citation impact, probability of getting breakthroughs, and probability of getting particularly poor outcomes - in expected ways. In other words, having greater value for *team_experience_diversity* and *team_network_size* is positively associated with average citation impact and with the likelihood of particularly high-impact innovations, and negatively associated with the probability of particularly low-impact innovations.

The third and most crucial step, reported in Table 5, involves checking whether the absolute size of the estimated effect of *team_size* (the independent variable) on each of the dependent variables decreases significantly once the mediators are included in the respective regression. This step is shown for the three dependent variables—overall citation impact, likelihood of being among the top 5% in citation impact, and likelihood of getting zero citations in columns (1) and (2), columns (4) and (5), and columns (7) and (8), respectively. The size of the mediation effect is found to be substantial. For example, once the mediator variables are introduced, the difference between expected citation impact for lone inventors versus four-person teams falls from 23% to 14%. Likewise, the difference between lone inventors and four-person teams on the probability of breakthroughs falls from 64% to 34%. Similarly, the magnitude of the difference between lone inventors and four-person teams on the probability of zero citation outcomes falls from 28% to 22%. Overall, all three sets of regressions demonstrate the effect operating through both of the proposed mediators: once the proposed mediators are controlled for, the observed benefits of collaboration decrease significantly in all three cases.

A closer look at team processes: Implications for the extent and breadth of technological search

The results presented so far appear to be consistent with our evolutionary argument of how collaboration changes the processes of invention. However, if the above arguments are correct, we should also observe that collaboration leads to characteristics of the inventive product being more like those commonly associated with high-impact inventions. Restated in our mediation terminology, the effects of team size, team diversity, and social networks on the final impact should be mediated by the characteristics of the invention itself. To develop this idea, we analyze the effect that collaboration has in shaping the depth and breadth of technological search going into an invention. As summarized in Table 2, we use three variables to capture the search process. *Patent_references* (the number of citations a patent makes of previous patents) and *nonpatent_references* (the number of non-patent references a patent makes) help capture the extent of search within existing knowledge. *Breadth_of_search* (a Herfindahl-based measure of the extent

to which backward citations are dispersed across different technologies) helps capture the breadth of search across technological areas, and is formally defined for patent *i* as

$$1-\sum_{j} s_{ij}^{2},$$

where s_{ij} refers to the fraction of patents cited by patent *i* that belong to technology class *j* (Jaffe and Trajtenberg 2002).

In analysis omitted here in order to conserve space, we found all the above search-related variables to be associated positively with larger teams, suggesting that collaboration does indeed increase both the extent and the breadth of technological search. As with the team-level mediators examined earlier, we introduced these variables as additional mediators in the analysis for Table 5. The results for our three dependent variables—citation impact, likelihood of being among top 5% in impact, and likelihood of getting zero citations—are documented in column (3), column (6), and column (9), respectively. Three observations are worth making in comparing columns (2) and (3). First, as the positive and significant coefficients for *patent_references*, *non_patent_references*, and *breadth_of_search* indicate, both the extent and breadth of search have a positive effect on an invention's citation impact. Second, including these variables reduces the estimated coefficients for *team_experience_diversity* and *team_network_size*, confirming our intuition that teams with more diverse experience and larger networks are more likely to lead to high-impact outcomes at least in part due to such teams' ability to conduct more extensive and wider search. Finally, the direct estimate for *team_size* itself also decreases, suggesting that the two team-level mediators we used above are not exhaustive in explaining the search-related benefits associated with larger teams.

The search-related findings for extreme outcomes are also interesting. Notice that breadth of search seems to be relatively more valuable for achieving breakthroughs (column (6)) than for avoiding particularly bad outcomes (column (9)), while the extent of search of patented knowledge seems more useful for avoiding

particularly bad outcomes than for achieving breakthroughs. Comparing the difference in coefficients between columns (5) and (6) with that between columns (8) and (9) also reveals some interesting differences. Looking at likelihood of high impact, it appears that better search explains gains from experience diversity to a greater extent than it explains gains from larger networks. On the other hand, looking at whether a patent is among those with no citation impact at all, better search explains gains from larger networks at least as much as it explains gains from experience diversity. Given the low statistical significance of these differences between coefficients, these findings are only suggestive, but worth exploring further in future research.

Quantile regression analysis

Having separately analyzed the effect of collaboration on the mean number of citations and on the tails of the distribution, we employ "quantile regression" methodology to examine the entire distribution of outcomes (Koenker and Bassett 1978; for its first application in the study of creativity, see Girotra, Terwiesch, and Ulrich 2007). Unlike classical regression, which relates the mean of a dependent variable to the explanatory variables, quantile regression estimates how the relationship varies for different percentiles of the data. As before, our dependent variable is the number of citations received by a patent. Table 6 reports the results of the quantile models. The results clearly demonstrate that larger teams dominate smaller teams and the lone inventor across all percentiles. The effects between adjacent quantiles are significantly different, demonstrating more than a simple mean shift.[10] Further, the marginal gains from having larger teams are significantly larger for the higher quantiles. Figure 3 graphically depicts regression estimates for *ln_team_size* for different quantiles, along with what the analogous estimates turn out to be if the two mediator variables, *team_experience_diversity* and *team_network_size*, are also included in the above model. As expected, we see a significant decrease in the coefficients for *ln_team_size* across all quantiles once the mediators have been included.

[10] If the benefit of collaboration were only in the form of a mean shift, we would still observe positive estimates across all the quantiles, but these estimates would no longer be significantly different from one another.

DISCUSSION

Building on a model of evolutionary search with its three classic phases of variation, selection, and retention, we argued that collaboration (1) enables more careful and rigorous selection of the best ideas and (2) increases the combinatorial opportunities for novelty. Collaborative teams are therefore less likely to invent useless inventions and more likely to invent breakthroughs. We found supporting correlations in an archival dataset of patent data; collaboration appears to trim the lower tail of the distribution of citations to a patent and enhance the number of far-right outliers. The technical diversity of the team and the team's external collaborative network appear to at least partially mediate both effects.

While our archival database enables this study by furnishing enough data to model extremely rare events, it also imposes a number of limitations. First, as is typical of almost studies on collaboration that employ archival data, our results are based on a cross-sectional comparison of pre-existing (and successfully patenting) teams. Compared to laboratory teams, analysis of pre-existing and real-world teams provides greater external validity to our results, but at the cost of not having a random assignment of individuals to different teams (and indeed, to individual vs. collaborative effort). Since we do not have an exact model of how individuals get assigned to specific teams, we cannot be sure if the same inventor would have behaved according to our predictions in different collaborative situations. Thus, our results do not enable causal inference. As a robustness check, we did ensure that our main findings are robust to using panel-data models with inventor fixed effects, at least mitigating concerns that unobserved time-invariant characteristics of inventors (e.g., better inventors being more sought after for collaboration) might be driving the results. Admittedly, to the extent that unobserved individual characteristics might change over time, or that the expectations for the outcome might itself influence the likelihood of collaboration, a fixed-effects model does not perfectly resolve the issue of collaboration being endogenous. A natural experiment or an instrumental variable approach would have been ideal to address these concerns, but neither was found in our setting.

Another concern, which again applies not just to our study but to most existing research on collaboration, is that absence of fine-grained cost data prevents us from making conclusive statements about the *net* value of collaboration. In other words, would a number of inventors be more likely to invent a breakthrough if they collaborated from the start, of if they worked alone and then pooled their efforts at the end? (Or following Girota, Terwiesch, and Ulrich, 2007, at what point in the process should inventors start collaborating?) It is probable, for example, that collaboration imposes more administrative costs than individuals working alone (as a trivial but obvious cost, only one person can talk at a time during meetings). In addition, collaborative projects might have preferential access to financial resources like R&D funds or research grants. Finally, the input of *person-hours* probably correlates with the size of the team that came up with the patent. High-impact patents arise from novel insights in addition to "brute work", however, and such insights almost surely require much more than just raw labor. As a preliminary attempt at mitigating concerns about comparing patents with different associated costs, all our models instead control for the number of claims, with the expectation than patents requiring more input costs are also likely to end up with a greater number of claims. Research into the costs - and opportunity costs - of collaboration strikes us as an important topic for future work.

Another concern with our study could be whether using patent data produces selection bias in the results. In relying on archival data, we only observe creative efforts that resulted in a patent. We are therefore not able to do justice to the large social-psychology literature that documents many detailed nuances of group processes, demographics, and structure (Paulus and Nijstad 2003) that surely influence whether an effort results in a patent. Against these qualifications, we should highlight the unique strengths of our study – a large scale examination of the entire distribution of creative outcomes, wherein we find strong and robust results as well as significant mediation effects consistent with theory. Even if a similarly powerful lab study could be performed (and require hundreds or even thousands of subjects, due to the rarity of creative outliers), such a lab study would still be unable to provide the long-term and social analysis of breakthroughs that occurs in the real world.

At a minimum, we hope this article prompts a rethink of current approaches to researching collaborative creativity. We hope that theorists will increase their metaphorical "degrees of freedom," that rather than focusing only on the mean and/or the variance of a distribution, they begin thinking about the different effects a variable may have on the lower tail, the mean, and the upper tail. For example, the conflicting results for lone inventor variance (Dahlin, Taylor, and Fichman 2004 and Fleming 2007 versus Taylor and Greve 2006 and Girotra, Terwiesch, and Ulrich 2007) were probably driven by an increase in lone inventor failures, rather than by a symmetric increase in failures and breakthroughs. But the variance-based models in these conflicting studies implicitly assumed symmetry of the tails and therefore masked the asymmetric influence of collaboration (though Girotra, Terwiesch, and Ulrich 2007 estimate a quantile model). While focusing on the entire distribution of outcomes will obviously increase the complexity of research, it should also provide more accurate models and more nuanced predictions of the phenomenon of creativity. Our results also highlight the need for empirical and methodological caution in general, wherein failure to account for different influences of explanatory variables in different parts of the creative distribution may provide only limited understanding of the phenomenon, particularly if the quality or impact of ideas varies greatly.

Though not the focus of our paper, our results also appear to have implications for theories of learning curves. Much of what we know about the value of experience derives from settings such as manufacturing, where the emphasis is on repeating routine tasks (Wright 1936). Our results indicate that experience might have much less benefit in creative contexts and that what benefit it does have might accrue from mechanisms very different from those at work in routine activities. For example, the estimate for *team_patents* turns negative and significant once the mediators have been introduced in column (3) or column (6) of Table 5. This suggests that benefits from a team's combined experience at innovating operate largely through associated exposure to diverse technologies and access to a wider community of researchers through the team members' interpersonal networks. If these factors were to be held constant,

mere scale of a team's combined experience actually appears to be detrimental for performance. This raises a question as to whether the benefits from "learning curves", typically examined in the manufacturing context, really apply to creative contexts. This opens up another area for future research.

Further progress using evolutionary analogies for creativity will require a combination of laboratory and archival methods. An evolutionary model of creativity is extremely difficult to study in its entirety because it unfolds over time and space (where space can be defined quite broadly, including social, geographical, technical, or organizational space) and over levels of analysis (psychological; social-psychological; and social, economic, or industrial). Some methods, such as laboratory experiments, are very well suited to understanding the earlier stages of the process. With lab experiments, the subjects, sequence of psychological and/or social iteration and other inputs to the process can be carefully controlled and randomized. But lab experiments cannot observe the social stage of retention by a larger community. Other methods, such as econometric analyses of large-sample archival data, are better suited to understanding the acceptance and success of ideas, but are less well suited to understanding the underlying processes of generation, because they only observe an idea after it has been selected for publication. Since lab methods and archival studies appear complementary, the research streams should benefit greatly from awareness and cross-fertilization. Furthermore, given the intuitive difficulty of conceptualizing outliers and higher moments, researchers should also incorporate formal models (e.g., Dahan and Mendelson 2001; Girotra, Terwiesch, and Ulrich 2007) and econometric innovations (e.g., Koenker and Bassett 1978) into their research.

CONCLUSION

Are people more creative – and more likely to invent breakthroughs - when they work alone or together? While recent research on the topic has tended to favor collaboration, other research has continued to laud the solitary inventor. Recent inquiry has attempted to resolve this controversy by investigating how collaboration influences the variability of creative outcomes, but the results have been contradictory. To resolve the controversy, we consider the entire distribution of creative outcomes. We argue that collaboration trims the bottom of the distribution, due to better selection processes; solitary inventors are less effective than groups at culling the bad ideas. Collaboration also increases the number of successful outliers because the diversity of groups enables more novel combinations. Greater diversity within a group and larger external networks beyond a group appear to mediate these processes. Raw data and quantile estimations from patent citations support these arguments. While the data did not afford natural experiments or instruments for causality, the results are robust and hold for fixed-effects analyses.

The results do not support the accounts of the heroic lone inventor; the story appears to be a myth, at least when considering the ultimate success of an invention. While we agree with Steinbeck that the original insight occurs within an individual brain, it also appears that creative processes ultimately benefit greatly from collaboration. The stochastic dominance of collaborative patents over individual patents is particularly telling; at every point in the distribution of creative outcomes, collaboration appears to increase the probability of a desired result. Still, the myth might have been (or may still be) true. It may be that the lone inventors of the 19th century were truly heroic, but that 20th century changes in technology increased the advantages of collaboration. On the other hand, it may be that, even today, the opportunity costs of inventing together may be greater than those of simply working alone and then summing the individual efforts. In any case, the larger goal of this paper was to motivate a re-examination of how we study creativity. However future researchers think about and study these issues, we hope this work highlights the importance of studying the asymmetric moments of collaborative creativity.

REFERENCES

Albert, M. B., Narin, F., Avery, D., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*, 251-259.

Alcacer, J., & Gittelman, M. (2006). Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations. *Review of Economics and Statistics* 88(4), 774-779.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology, 51*(6), 1173-1182.

Basalla, G. (1988). *The evolution of technology*. Cambridge, MA: Cambridge University Press.

Brefka, P. (2006). Personal interview. Boston, MA, August 28, 2006.

Campbell, D. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*, 380-400.

Csikszentmihalyi, M. (1999). *Creativity*. New York: Harper Collins.

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

Dahlin, K., Taylor, M., & Fichman, M. (2004). Today's Edisons or weekend hobbyists: Technical merit and success of inventions by independent inventors. *Research Policy, 33*(8), 1167-1183.

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

Evans, H. (2005, June). The eureka myth. *Harvard Business Review*.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science, 47*(1), 117-132.

Fleming, L. (2002). Finding the organizational sources of breakthrough invention: The story of Hewlett Packard's ink jet invention. *Industrial and Corporate Change, 11*(5), 1059-1084.

Fleming, L. (2007). Lone inventors as sources of breakthroughs: Myth or reality? Working paper, Harvard Business School, Boston.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly, 52*(3), 443-475.

Gilfillan, S. (1935). *The sociology of invention*. Cambridge, MA: MIT Press. Reprinted 1970.

Girotra K., Terwiesch, C., & Ulrich, K. (2007). Idea generation and the quality of the best idea. Working Paper, INSEAD, Fontainbleau.

Hall, B., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics, 1*, 16-38.

Harhoff, D., Narin, F., Scherer, F., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics, 81*(3), 511-515.

Hughes, T. (2004). *American genesis: A century of invention and technological enthusiasm, 1870-1970*. Chicago: Chicago University Press.

Jaffe, A., & Trajtenberg, M. (2002). *Patents, citations and innovations: A window on the knowledge economy.* Cambridge, MA: MIT Press.

Kenney, S. (2006). Personal interview. Cambridge, MA, Sept. 1, 2006.

King, G. (1989). Event count models for international relations: Generalizations and applications. *International Studies Quarterly, 33,* 123-147.

Koenker, R., & Bassett, G. W. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Levin, R., Klevorick, A., Nelson, R., & Winter, S. (1987). Appropriating the returns from industrial research and development: Comments and discussion. *Brookings Papers on Economic Activity*, *3*, 783-831.

March, J. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2,* 71-87.

McFadyen, M., & Cannella., A. (2004). Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. *Academy of Management Journal*, *47*(5), 735-746.

Mokyr, J. (1990). *The lever of riches: Technological creativity and economic progress*. New York: Oxford University Press.

Morison, E. (1966). *Men, machines, and modern times*. Cambridge, MA: MIT Press.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology, 12,* 3-23.

Paulus, P., & Brown, V. (2003). Enhancing ideational creativity in groups. In P. Paulus and B. Nijstad (Eds.), *Group creativity: Innovation through collaboration.* New York: Oxford University Press.

Paulus, P. & Nijstad, B. (2003). Group creativity: An introduction. In P. Paulus and B. Nijstad (Eds.), *Group creativity: Innovation through collaboration.* New York: Oxford University Press.

Ratliff, E. (2006). Me translate pretty one day. *Wired, 14*(12), 210-213.

Romer, P. (1993, September 11). Ideas and things: The concept of production is being re-tooled. *The Economist*, 70-72.

Runco, M. (1995). Creativity need not be social. In A. Montuori and R. Purser (Eds.), *Social creativity* (pp. 237-264). Creskill, NJ: Hampton Press.

Runco, M., & Smith, W. (1991). Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences, 13,* 295-302.

Schumpeter, J. A. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.

Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy*. New York: Harper.

Schwartz, E. (2002). *The last lone inventor: A tale of genius, deceit, and the birth of television.* New York: Harper Collins.

Schwartz, E. (2004). *Juice: The creative fuel that drives world-class inventors*. Boston: Harvard Business School Publishing.

Simonton, D. (1985). Quality, quantity, and age: The careers of 10 distinguished psychologists. *International Journal of Ageing and Human Development, 21,* 241-254.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. New York: Oxford University Press.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science, 51*(5), 756–770.

Smith, A. (1766, reprinted 1982). *Lectures in jurisprudence*. Indianapolis, IN: Liberty Fund.

Steinbeck, J. (1952). *East of Eden*. New York: Putnam Penguin.

Sutton, R., & Hargadon, A. (1996). Brainstorming groups in context: Effectiveness in a product design firm. *Administration Science Quarterly, 41*(4), 685-718.

Taylor, A., & Greve, H. (2006). Superman or the Fantastic Four? Knowledge combination and experience in innovative teams. *Academy of Management Journal, 49*(4), 723-740.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of. Economics, 21,* 172-187.

Trajtenberg, M., Shiff, G., & Melamed, R. (2006). The "names game": Harnessing inventors' patent data for economic research. NBER Working Paper No. 12479.

Tushman, M., & Anderson, R. (1986). Technological discontinuities and organizational environments. *Administration Science Quarterly, 31,* 439-465.

Weitzman, M. (1996, May). Hybridizing growth theory. *Proceedings of the American Economics Association,*207-212.

Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of Aeronautical Sciences, 3*(4), 122-128.

Wuchty, S., Jones, B., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Table 1. Citation impact for patents from teams of different sizes

	Observations	Mean	Standard Deviation	Coefficient of Variation	5th percentile	10th percentile	Median	90th percentile	95th percentile	99th percentile
Lone Inventor	269,823	9.47	14.30	1.51	0	1	6	21	31	66
Team size = 2	146,677	10.97	15.71	1.43	0	1	6	25	36	75
Team size = 3	73,515	12.32	17.83	1.45	0	1	7	29	42	83
Team size = 4	31,666	13.47	19.96	1.48	0	1	7	31	46	94
Team size = 5	12,923	14.93	21.79	1.46	0	1	8	35	52	104
Team size >= 6	11,510	17.40	28.74	1.65	0	1	9	41	60	118
Overall	546,114	10.79	16.27	1.51	0	1	6	24	36	76

Table 2. Variable definitions and summary statistics

		Mean	Std Dev	Min	Max
Dependent variable					
crec	Number of citations received by the patent (until 2004)	10.79	16.27	0	1,249
Explanatory and control variables					
team_size	Size of the inventing team	1.94	1.31	1	34
unassigned	Indicator variable that is 1 if and only if patent has no assignee	0.22	0.42	0	1
university	Indicator variable that is 1 if and only if patent is assigned to a university	0.03	0.17	0	1
claims	Number of claims made by the patent	14.67	11.81	1	320
inventor_age	Mean number of active years for this team's inventors	4.36	4.90	0	35
team_patents	Number of unique past patents involving anyone from this team	9.33	23.24	0	694
Team-level mediators for effect of team size on citations received by patent					
team_experience_diversity	Number of technology classes any team inventor has patented in before	6.15	9.39	0	234
team_network_size	Number of inventors at distance <=2 in the team's collaborative network	11.57	29.62	0	991
Variables directly capturing search behavior of the patenting team					
patent_references	Number of backward citations that the patent makes to other patents	10.68	11.92	0	745
nonpatent_references	Number of non-patent references made by the patent	2.30	7.48	0	100
breadth_of_search	One minus the Herfindahl of technological spread of backward citations	0.41	0.28	0	0.95

Table 3. Correlation matrix among variables

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	crec	1.000											
(2)	team_size	0.110	1.000										
(3)	unassigned	-0.086	-0.282	1.000									
(4)	university	0.020	0.051	-0.091	1.000								
(5)	claims	0.150	0.123	-0.110	0.039	1.000							
(6)	inventor_age	-0.024	-0.006	-0.144	-0.020	0.042	1.000						
(7)	team_patents	0.000	0.224	-0.099	-0.014	0.033	0.384	1.000					
(8)	team_experience_diversity	0.041	0.276	-0.154	-0.001	0.054	0.493	0.812	1.000				
(9)	team_network_size	0.044	0.362	-0.177	-0.014	0.049	0.194	0.506	0.490	1.000			
(10)	patent_references	0.086	0.083	-0.036	-0.032	0.176	0.050	0.053	0.055	0.047	1.000		
(11)	nonpatent_references	0.063	0.145	-0.104	0.215	0.120	0.018	0.043	0.052	0.089	0.188	1.000	
(12)	breadth_of_search	0.069	0.075	-0.082	0.000	0.075	0.022	0.024	0.087	0.050	0.278	0.070	1.000

Table 4. Regression analysis for how collaboration affects citation impact

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*crec*	*crec*	*Top 5% in citation*	*Top 1% in citation*	*No citations*	*<= 1 citations*
Regression model:	*Negative binomial (clustered SE)*	*Negative binomial (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*
ln_team_size	0.246***	0.149***	0.356***	0.361***	-0.240***	-0.249***
	(0.0062)	(0.0085)	(0.021)	(0.057)	(0.015)	(0.011)
unassigned		-0.162***	-0.518***	-0.651***	0.143***	0.138***
		(0.0067)	(0.026)	(0.063)	(0.016)	(0.012)
university		0.106***	0.331***	0.362***	-0.134***	-0.105***
		(0.021)	(0.048)	(0.091)	(0.038)	(0.029)
ln_claims		0.228***	0.544***	0.723***	-0.358***	-0.343***
		(0.0032)	(0.011)	(0.025)	(0.0083)	(0.0060)
ln_inventor_age		-0.0537***	-0.139***	-0.208***	0.00540	0.0116
		(0.0064)	(0.017)	(0.041)	(0.012)	(0.0088)
ln_team_patents		0.0608***	0.158***	0.254***	0.0370***	0.00854
		(0.0075)	(0.016)	(0.038)	(0.012)	(0.0091)
Year fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*	*Included*
Technology fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*	*Included*
Observations	546,114	539,365	539,365	539,365	539,365	539,365
chi2	26130	37029	5158	2033	12292	19170
df_m	45	50	50	50	50	50
ll	-1822191	-1793736	-107341	-30458	-122887	-214622

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5. Test for the mediation effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable:	*crec*	*crec*	*crec*	*Patent top 5% in citation*	*Patent top 5% in citation*	*Patent top 5% in citation*	*Patent gets no citations*	*Patent gets no citations*	*Patent gets no citations*
Regression model:	*Negative binomial (clustered SE)*	*Negative binomial (clustered SE)*	*Negative binomial (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*	*Logistic (clustered SE)*
ln_team_size	0.149***	0.109***	0.0929***	0.356***	0.253***	0.212***	-0.240***	-0.193***	-0.177***
	(0.0085)	(0.0083)	(0.0076)	(0.021)	(0.021)	(0.020)	(0.015)	(0.015)	(0.015)
unassigned	-0.162***	-0.132***	-0.114***	-0.518***	-0.442***	-0.398***	0.143***	0.110***	0.0944***
	(0.0067)	(0.0071)	(0.0067)	(0.026)	(0.026)	(0.026)	(0.016)	(0.017)	(0.017)
university	0.106***	0.106***	0.0683***	0.331***	0.335***	0.218***	-0.134***	-0.139***	-0.171***
	(0.021)	(0.021)	(0.021)	(0.048)	(0.048)	(0.050)	(0.038)	(0.038)	(0.041)
ln_claims	0.228***	0.228***	0.195***	0.544***	0.546***	0.459***	-0.358***	-0.357***	-0.308***
	(0.0032)	(0.0032)	(0.0030)	(0.011)	(0.011)	(0.011)	(0.0083)	(0.0083)	(0.0087)
ln_inventor_age	-0.0537***	-0.0689***	-0.0643***	-0.139***	-0.182***	-0.169***	0.00540	0.0340**	0.0367***
	(0.0064)	(0.0082)	(0.0073)	(0.017)	(0.020)	(0.020)	(0.012)	(0.013)	(0.014)
ln_team_patents	0.0608***	-0.0961***	-0.0822***	0.158***	-0.226***	-0.194***	0.0370***	0.205***	0.198***
	(0.0075)	(0.011)	(0.0099)	(0.016)	(0.025)	(0.025)	(0.012)	(0.020)	(0.020)
ln_team_experience_diversity		0.159***	0.135***		0.398***	0.337***		-0.195***	-0.185***
		(0.0074)	(0.0072)		(0.022)	(0.022)		(0.019)	(0.019)
ln_team_network_size		0.0590***	0.0566***		0.145***	0.143***		-0.0590***	-0.0559***
		(0.0046)	(0.0042)		(0.011)	(0.010)		(0.0082)	(0.0088)
ln_patent_references			0.110***			0.181***			-0.319***
			(0.0040)			(0.013)			(0.012)
ln_nonpatent_references			0.102***			0.231***			-0.0781***
			(0.0035)			(0.0098)			(0.010)
breadth_of_search			0.142***			0.461***			-0.0412
			(0.0085)			(0.030)			(0.026)
Year Dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Technology Dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included
Observations	539,365	539,365	525,252	539,365	539,365	525,252	539,365	539,365	525,252
chi2	37029	37963	41249	5158	5819	6988	12292	12326	12912
df_m	50	52	55	50	52	55	50	52	55
ll	-1793736	-1791827	-1745213	-107341	-106704	-103557	-122887	-122741	-115810

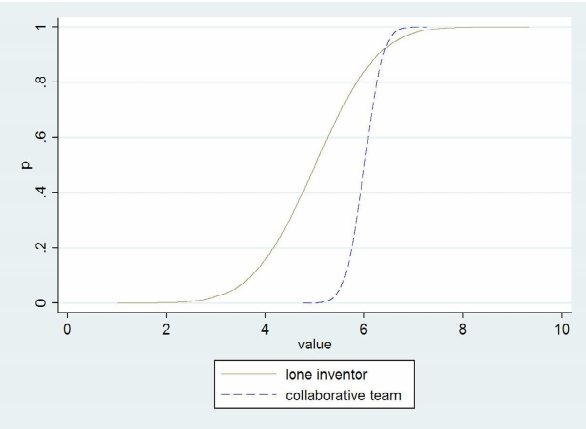
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Quantile regressions for effect of collaboration on the distribution of outcomes

	q10	q30	q50	q70	q90	q95	q99
ln_team_size	0.279***	0.647***	1.037***	1.685***	3.429***	4.859***	7.303***
	(0.012)	(0.015)	(0.029)	(0.053)	(0.11)	(0.21)	(0.74)
unassigned	-0.140***	-0.361***	-0.625***	-1.082***	-2.535***	-3.525***	-7.201***
	(0.011)	(0.019)	(0.027)	(0.044)	(0.12)	(0.14)	(0.48)
university	0.0806***	0.285***	0.511***	0.972***	3.220***	5.469***	6.816***
	(0.024)	(0.051)	(0.075)	(0.12)	(0.43)	(0.68)	(2.01)
ln_claims	0.353***	0.808***	1.282***	2.007***	3.754***	5.128***	9.416***
	(0.0080)	(0.011)	(0.015)	(0.026)	(0.074)	(0.086)	(0.29)
ln_inventor_age	-0.0260***	-0.105***	-0.217***	-0.387***	-1.039***	-1.627***	-3.626***
	(0.0093)	(0.013)	(0.021)	(0.032)	(0.070)	(0.14)	(0.28)
ln_team_patents	0.00338	0.0700***	0.201***	0.403***	1.144***	1.963***	4.921***
	(0.0069)	(0.010)	(0.019)	(0.032)	(0.080)	(0.17)	(0.37)
Year Dummies	Included	Included	Included	Included	Included	Included	Included
Technology Dummies	Included	Included	Included	Included	Included	Included	Included
Observations	539,365	539,365	539,365	539,365	539,365	539,365	539,365
Pseudo R2	0.016	0.040	0.059	0.087	0.140	0.168	0.213

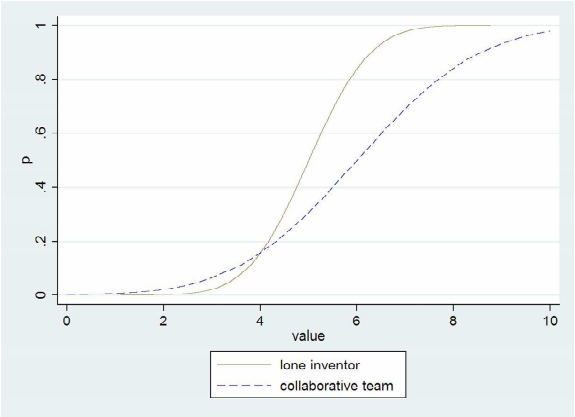
Figure 1: Hypothetical cumulative distribution functions

Curves demonstrate why studying variance alone could be insufficient to understand how collaboration affects probability of extreme outcomes.



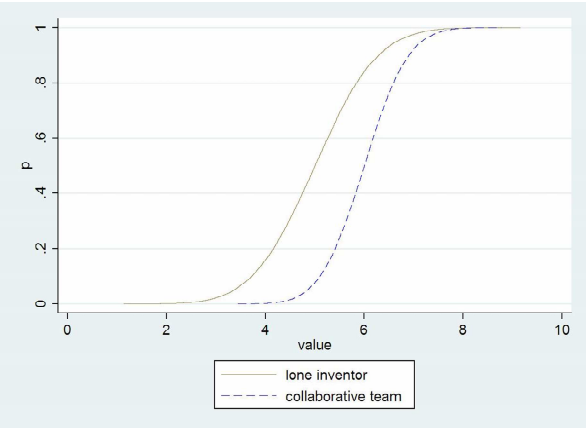
Scenario (1): *lower* variance from collaboration

Collaboration *decreases* probability of both particularly good and particularly poor outcomes



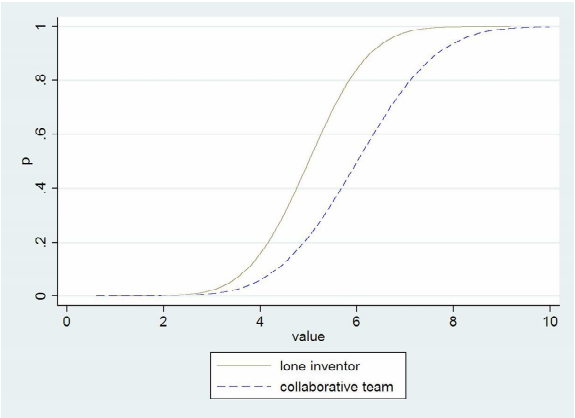
Scenario (2): *greater* variance from collaboration

Collaboration *increases* probability of both particularly good and particularly poor outcomes



Scenario (3): *lower* variance from collaboration

But collaboration now *decreases* probability of particularly poor outcomes and *increases* probability of particularly good outcomes



Scenario (4): *greater* variance from collaboration

But collaboration still *decreases* probability of particularly poor outcomes and *increases* probability of particularly good outcomes

Figure 2: Cumulative Distribution Functions of observed citation outcomes for innovations from teams of different sizes

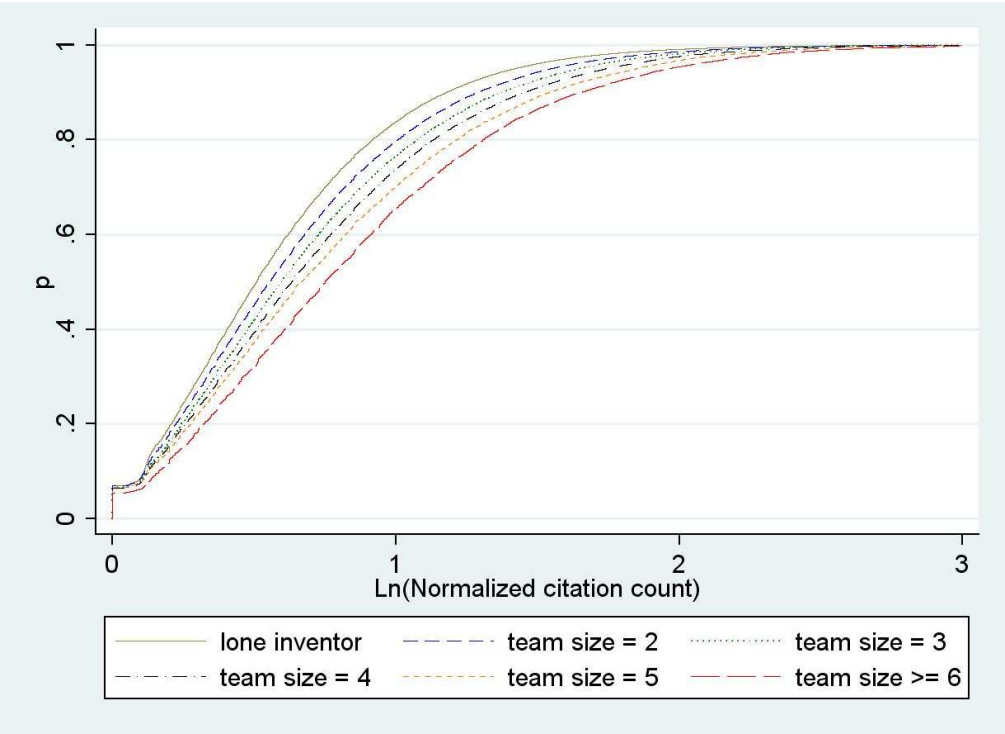
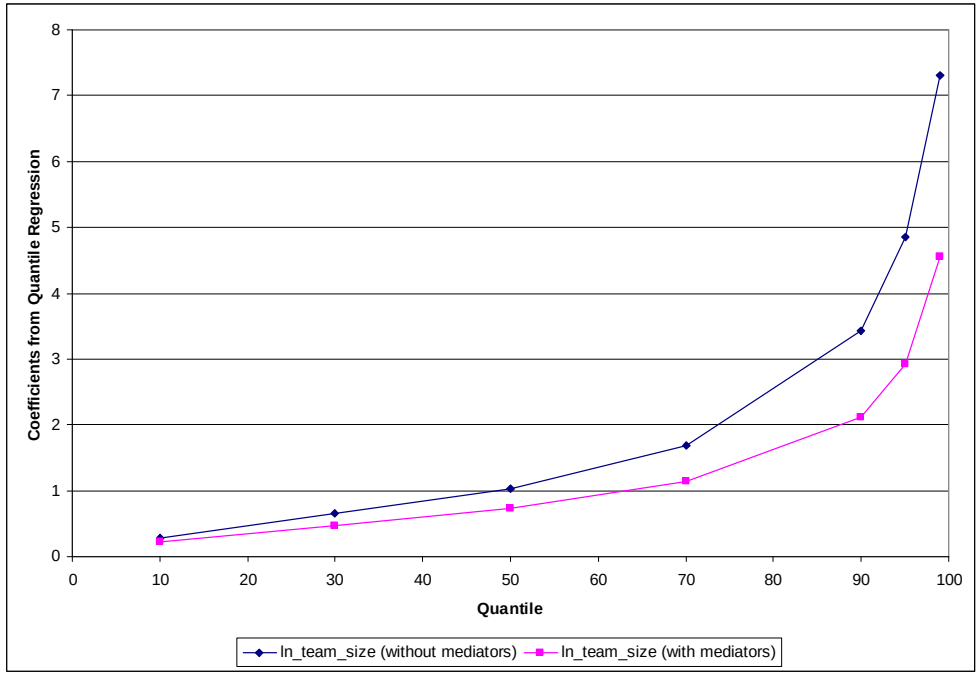


Figure 3: Coefficients for *ln_team_size* in quantile regressions (without vs. with mediators included)



Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

www.insead.edu

Printed by INSEAD

INSEAD

The Business School
for the World®