MANAGEMENT SCIENCE
Vol. 56, No. 1, January 2010, pp. 41–56
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5601 | 0041

informs®

DOI 10.1287/mnsc.1090.1072

Lone Inventors as Sources of Breakthroughs: Myth or Reality?

Jasjit Singh
INSEAD, Singapore 138676, Singapore, jasjit.singh@insead.edu

Lee Fleming
Harvard Business School, Harvard University, Boston, Massachusetts 02163, lfleming@hbs.edu

Are lone inventors more or less likely to invent breakthroughs? Recent research has attempted to resolve this question by considering the variance of creative outcome distributions. It has implicitly assumed a symmetric thickening or thinning of both tails, i.e., that a greater probability of breakthroughs comes at the cost of a greater probability of failures. In contrast, we propose that collaboration can have opposite effects at the two extremes: it reduces the probability of very poor outcomes—because of more rigorous selection processes—while simultaneously increasing the probability of extremely successful outcomes—because of greater recombinant opportunity in creative search. Analysis of over half a million patented inventions supports these arguments: Individuals working alone, especially those without affiliation to organizations, are less likely to achieve breakthroughs and more likely to invent particularly poor outcomes. Quantile regressions demonstrate that the effect is more than an upward mean shift. We find partial mediation of the effect of collaboration on extreme outcomes by the diversity of technical experience of team members and by the size of team members' external collaboration networks. Supporting our meta-argument for the importance of examining each tail of the distribution separately, experience diversity helps trim poor outcomes significantly more than it helps create breakthroughs, relative to the effect of external networks.

Key words: creativity; collaboration; invention; innovation; teams; quantile; diversity; networks
History: Received December 25, 2006; accepted July 21, 2009, by Christoph Loch, technological innovation product development, and entrepreneurship. Published online in *Articles in Advance* October 16, 2009.

1. Introduction

> 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. (Steinbeck 1952, pp. 130–131)

Are lone inventors more likely to generate breakthroughs—or is that just a myth? Nobel-prizewinning author John Steinbeck offers an eloquent testimonial to the creative abilities of the individual. He is not alone; writers, historians, and inventors have long championed the "lone inventor" over the group in the realm of creativity and, in particular, in the invention of breakthroughs (Schumpeter 1934; Mokyr 1990, p. 295; Hughes 2004, p. 53). Many creativity researchers have supported these arguments by elaborating the problems of creative teams, including idea blocking, communication difficulties, and interpersonal tensions (Diehl and Stroebe 1987, Mullen et al. 1991, Dougherty 1992, Runco 1999, Paulus and Brown 2003). Even proponents of teamwork acknowledge that research on the benefits of collaborative creativity remains "somewhat weak" (Paulus and Nijstad 2003, p. 4). Such modesty notwithstanding, proponents of collaboration have begun to successfully question the myth of the lone inventor. They have documented many disadvantages of individual effort (Sutton and Hargadon 1996, Paulus and Nijstad 2003, Perry-Smith and Shalley 2003, McFadyen and Cannella 2004) and the almost ubiquitous trend toward collaboration in science research (Wuchty et al. 2007). Much disagreement remains (Paulus and Nijstad 2003), however, and in particular, whether lone inventors are more or less likely to invent breakthroughs.

The goal of this paper is to enlarge the solution space for these debates on the value of collaboration, particularly with regard to the sources of breakthroughs and the processes by which they are conceived and developed. We propose that research on collaboration and creativity should place more emphasis on theorizing about and examining the entire distribution of creative outcomes. This is substantively important because such distributions tend to be extremely skewed, with most inventions being of little practical significance and a minute few being disproportionately impactful. We develop this argument in the context of lone inventors, where a lone inventor is socially isolated and either does not work with coinventors in a team, does not work for an organization, or both. Although most studies on collaborative creativity focus on the issue of individuals versus collaborative teams, and a few studies focus on "garage inventors" who do not work within an organization, very few studies consider the two contexts simultaneously. We propose that many of the arguments linking collaborative teams with idea generation also generalize to a comparison of ideas generated within versus outside an organization.

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. Motivated by an interest in particularly successful outcomes, or breakthroughs, recent work has empirically modeled the second moment or variance of creative outcome distributions (Dahlin et al. 2004, Taylor and Greve 2006, Girotra et al. 2007, Fleming 2007). Greater variance can increase the chances of a breakthrough because the associated increase in the mass in both tails implies a greater number of breakthrough outliers (Campbell 1960, Simonton 1999). These ideas parallel March's (1991) argument that structures, activities, and mechanisms that increase mean performance through exploitation might be quite different from those that achieve breakthroughs through exploration by increasing the variance in performance.

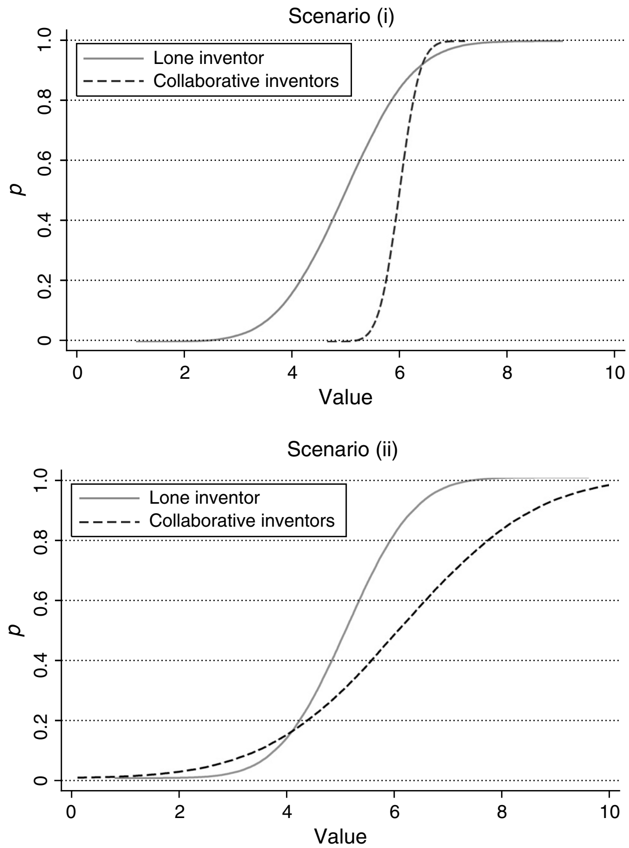
Unfortunately, the cumulative evidence on collaboration and breakthroughs remains ambiguous and indeed often conflicting. Dahlin et al. (2004) demonstrate that independent or garage inventors (those who do not work for an organization) are overrepresented in the tails of creative distributions. Consistent with this evidence, Fleming (2007) uses mean-variance decomposition models (King 1989) to show a lower average and greater dispersion of creative outcomes by individuals who work alone. In contrast to these results, Taylor and Greve (2006) demonstrate that collaboration in teams leads to higher variance in deviation from a normalized mean measure. Girotra et al. (2007) adopt an experimental approach and likewise demonstrate higher variance in outcomes generated by teams.

An examination of just the variance of outcomes, however, does not present the complete picture of how collaboration affects the distribution of creative outcomes. The hypothetical cumulative distribution functions shown in Figures 1(a) and 1(b) illustrate different ways in which collaboration could affect the outcome distribution, even assuming that collaboration is beneficial on average (many would dispute the assumption; see Paulus and Nijstad 2003). The debate is often framed as whether the observed variance of outcomes is more in line with scenario (i) or (ii) in Figure 1(a): If individual inventors are associated with greater variance of outcomes, scenario (i) is implicitly assumed and collaboration is therefore judged as being less desirable for achieving breakthroughs. On the other hand, if individual inventors are associated with lower variance, then scenario (ii) is implicitly assumed and collaboration is considered more desirable. Note, however, that both scenarios in Figure 1(a) assume symmetry in how collaboration affects the extremes: Neither allows the possibility that increased likelihood of breakthroughs could come without a corresponding increase in likelihood of particularly bad outcomes. As scenarios (iii) and (iv) in Figure 1(b) illustrate, achieving greater variance is in reality neither necessary nor sufficient for ensuring greater likelihood of breakthroughs. Whereas lone inventors are associated with greater variance in scenario (iii) and lower variance in scenario (iv), they are worse at achieving breakthroughs in both these scenarios. In other words, the effects at the two tails need not involve a trade-off: collaboration can increase the likelihood of breakthroughs while simultaneously reducing the probability of particularly bad outcomes.[1]

Building on a stylized evolutionary model of creativity, we explore why lone inventors might be less likely to invent breakthroughs and more likely to invent failures. Supporting an argument that collaboration improves the sorting and identification of promising new ideas, we find that working as a part of a team or an organization, or both, trims the lower tail of the distribution of outcomes. On the other hand, and supporting an argument that collaboration enables more creative novelty, we find that team and/or organization affiliation increases the likelihood of creative outcomes toward 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). Consistent with an argument that a greater diversity of knowledge enables greater recombinant opportunity and more rigorous assessment of that opportunity, we find two partial

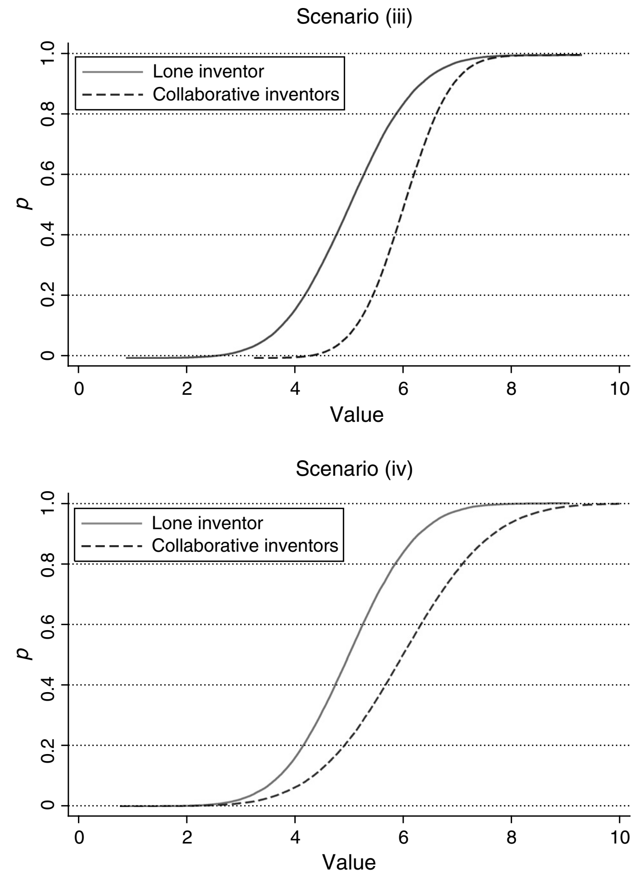
[1] These examples should not be taken literally. For ease of modeling, the illustrative scenarios in Figure 1 were generated by assuming a normal distribution where collaboration increases the mean but is allowed to increase or decrease the variance to different extents. In most distributions, larger variance *will* in the limit have a greater probability of the very extreme outcomes at both ends (Fleming 2007). However, the probability mass where this happens might be too trivial to be of economic significance. Furthermore, real-world outcomes need not obey a strictly normal—or for that matter even symmetric—distribution.

Figure 1(a) **Illustrative Scenarios Where Greater Variance and Greater Likelihood of Breakthroughs Occur Together**



Notes. Scenario (i): Lone inventors achieve greater overall variance, greater likelihood of breakthroughs, and greater likelihood of particularly poor outcomes. Scenario (ii): Lone inventors achieve lower overall variance, lower likelihood of breakthroughs, and lower likelihood of particularly poor outcomes.

Figure 1(b) **Illustrative Scenarios Where Greater Variance and Greater Likelihood of Breakthroughs Do Not Occur Together**



Notes. Scenario (iii): Lone inventors achieve greater overall variance, lower likelihood of breakthroughs, and greater likelihood of particularly poor outcomes. Scenario (iv): Lone inventors achieve lower overall variance, lower likelihood of breakthroughs, and greater likelihood of particularly poor outcomes.

mediators of collaboration: the diversity of past technological experiences of members in a collaborative team and the size of a team's external collaboration network. Supporting our meta-argument for the importance of examining each tail of the distribution separately, experience diversity helps trim poor outcomes significantly more than it helps create breakthroughs, relative to the effect of external networks.

2. Lone Inventors, Creativity, and Generation of Breakthroughs

Following many researchers (Campbell 1960, Romer 1993, Weitzman 1998, 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. The idea that novelty is a new combination is at least as old as Adam Smith (1766). Given thorough historical search, novel technologies can almost always be traced to combinations of prior technologies (Basalla 1988). Science, music, language, art, design, manufacturing, and many other forms of creative endeavor have been described similarly (Gilfillan 1935, Romer 1993, Weitzman 1998).

In the second or "selection" phase, inventors evaluate ideas to reject poor outcomes and identify the most promising novelties. The processes within the generation and selection 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 rare in today's interconnected world; in our data, the example corresponds most closely to an independent or garage inventor who works without a team. At the other extreme of very social 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). This archetypal example is increasingly common in today's world (Wuchty et al. 2007), and in our data it corresponds most closely to inventors who work in a team within an organization.

In the last or "retention" phase, members of a larger creative 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. Whereas 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. This argument is old; Hooker et al. (2003, p. 230), for example, argue, "To be creative, a variation must somehow be endorsed by the field.... Creativity involves social judgment." (See also Simonton 1999, Csikszentmihalyi 1999.) Creative individuals can incorporate their own prior work, but their influence will be quite 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 future inventions.

Independent of the idea's source (lone versus collaborative), we propose that collaboration improves the effectiveness of the selection phase because collaborative selection will be more rigorous than lone selection. 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 nonexperts (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, p. 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, p. 212). The dual inventors of the Hewlett Packard thermal inkjet printer were a prolific empirical tinkerer who generated prototypes and a very methodical engineer who explained, documented, and criticized (Fleming 2002).

Arguments similar to benefits from affiliation with teams can also be made for affiliation with organizations. We propose that a single independent inventor (the image here is of an antisocial individual working in his or her garage) will be more isolated than a single inventor that works within an organization. The assumption is that an affiliated inventor who does not collaborate will still enjoy more social interaction (among colleagues and technical experts) than an unaffiliated inventor. This assumption is consistent with perspectives that the ability to accumulate and leverage knowledge provides a key reason for the existence of firms (Nelson and Winter 1982, Grant 1996). Accordingly, firms can be seen as social communities that are a natural extension of teams when it comes to creation of new knowledge (Kogut and Zander 1992). Though there are surely exceptions of highly connected yet independent inventors, our argument depends on the typical independent inventor being more isolated than the typical affiliated inventor.

Because isolated inventors will lack multiple and (to varying degrees) uncorrelated filters, they will uncover fewer potential problems and hence develop more 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, giving our first hypothesis.

Hypothesis 1. *Lone inventors will invent a relatively greater proportion of low impact inventions than collaborative inventors.*

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 1998), we argue that collaboration should increase 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, such that the worse outcomes should 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 more efficiently generate *and* assess more potential options. Partly because of their faster and more efficient selection processes, such teams can spend more time on contriving radically novel combinations that have greater breakthrough potential. Immediate access to greater diversity also makes teams more efficient in generating radically new combinations in the first place. Thus, in addition to 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 the above reasons—greater diversity that enables greater recombinant opportunity and a greater volume of iterations—collaborative teams and affiliated inventors should generate more potential novelty at the breakthrough end of the distribution. This leads to our second hypothesis.

Hypothesis 2. *Lone inventors will invent a relatively lesser proportion of high impact inventions than collaborative inventors.*

Note that these predictions do not depend on the influence of collaboration or affiliation on the average outcome. Although we believe that collaboration should also influence the mean positively (consistent with McFadyen and Cannella 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 significantly different sizes on the two tails, in both logit models of extreme outcomes and quantile regressions). And though our data do not afford an exogenous experiment in lone versus collaborative invention, we develop and test additional observable implications of our theory with mediation analysis (Baron and Kenney 1986).

The arguments for the first two hypotheses depend heavily on the role of diversity. Following the logic of Hypothesis 1, we propose that diversity helps identify a poor outcome before it is fully developed, thus trimming the lower tail. Diverse teams will contain a greater variety of opinions and should be less vulnerable to groupthink when assessing the value of an idea (Janis 1972). Following the logic of Hypothesis 2, the diversity of team backgrounds should generate greater recombinant opportunity and potential applications, thus fattening the upper tail. Typically, as the size of a team increases, so should the aggregate diversity of experience within the team. These arguments imply that diversity mediates the value of collaboration, at both extremes. This leads to our third testable hypothesis.

Hypothesis 3. *The effect of collaboration on extreme inventive outcomes will be mediated by the diversity of experience of the team members.*

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 nonteam colleagues, they can suggest overlooked possibilities or problems.

Typically, as the size of a team increases, so should the size of the external network that the team can access. Therefore, similar to the aggregated experience diversity within the team, the size of the extended collaborative network should mediate the value of collaboration at both extremes. This implies the last prediction.

Hypothesis 4. *The effect of collaboration on extreme inventive outcomes will be mediated by the size of the extended social network of the team members.*

3. Data

We examine the link between lone inventors and the distribution of creative outcomes using U.S. Patent

Table 1(a) Raw Statistics Regarding Citation Impact of Patents from Individual Inventors vs. Teams

	Observations	Mean	Standard deviation	Coefficient of variation	5th percentile	10th percentile	Median	90th percentile	95th percentile	99th percentile
Individual inventor	260,438	9.49	14.30	1.51	0	1	6	21	31	66
Team size = 2	136,033	11.03	15.70	1.42	0	1	6	25	36	75
Team size = 3	67,588	12.52	17.96	1.43	0	1	7	29	42	83
Team size = 4	29,125	13.73	20.25	1.47	0	1	8	32	47	95
Team size = 5	11,906	15.30	22.18	1.45	0	1	8	36	53	105
Team size ≥ 6	10,726	17.77	29.30	1.65	0	1	9	41	61	122
Overall	515,816	10.84	16.31	1.51	0	1	6	24	36	76

Table 1(b) Raw Statistics Regarding Citation Impact of Unassigned vs. Assigned Patents

	Observations	Mean	Standard deviation	Coefficient of variation	5th percentile	10th percentile	Median	90th percentile	95th percentile	99th percentile
Unassigned	122,553	8.22	12.36	1.50	0	1	5	17	25	56
Assigned to firm	393,263	11.65	17.28	1.48	0	1	7	26	39	80
Overall	515,816	10.84	16.31	1.51	0	1	6	24	36	76

and Trademark Office (USPTO) patent data. These data are attractive for several reasons. First, they allow a systematic comparison of creative outcomes on both dimensions we are interested in: relative success of individuals versus teams as well as of individuals working independently versus within organizations. Second, future citations received by patents provide a systematic method of measuring impact in a way that is comparable across outcomes. Third, the longitudinal nature of the data allows us to examine a rich set of questions, including those requiring a historical account of past experiences of inventors in terms of both the kind of projects they have worked on before and the people they have collaborated with. Finally, being able to draw a large sample across a wide range of sectors increases the power of the statistical tests and makes findings more general. We constructed the data set from three sources: the USPTO itself, the National University of Singapore, and the National Bureau of Economic Research (Jaffe and Trajtenberg 2002, Chap. 13). We applied inventor-matching algorithms similar to those previously employed by Singh (2005, 2008), Trajtenberg et al. (2006), and Fleming et al. (2007) to create a reliable patent-inventor mapping. Assignee names and parent-subsidiary matching were corrected based on procedures described in Singh (2007).

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. Firms and individuals differ in their reliance on patents, often relying on alternative means of protecting their intellectual property (Levin et al. 1987). Nevertheless, conditional on a specific innovation being patented, citations to that patent help capture its overall economic, social, and technological success. The number of citations a patent receives has been shown to be correlated with several measures of 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 an organization's market value (Hall et al. 2005).[2]

We restrict our final sample to successful patents filed during the 10-year period 1986–1995, allowing sufficient historic as well as future time window for constructing our measures, such as prior experience of a patent's inventors and future citation impact of the patent. In constructing these measures, we use information from all USPTO patents granted during 1975–2004. However, for comparability during regression analysis, the actual sample analyzed (from 1986 to 1995) was further restricted only to patents arising from U.S.-based inventors.

A simple inspection of the data provides preliminary support for our arguments. As indicated in Table 1(a), patents resulting from teams appear to be associated with more citations than those from individual inventors, a benefit that appears to increase unambiguously with team size. Likewise, Table 1(b) shows patents assigned to organizations also receive more citations. The standard deviation of citation outcomes also increases with team and organization affiliation, though there is no obvious relationship with the coefficient of variation. The reported percentile

[2] This is also consistent with the view of the USPTO: "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" (Office of Technology Assessment and Forecast 1976, p. 167).

Figure 2(a) **Cumulative Distribution of Observed Impact of Patents from Individuals vs. Teams**

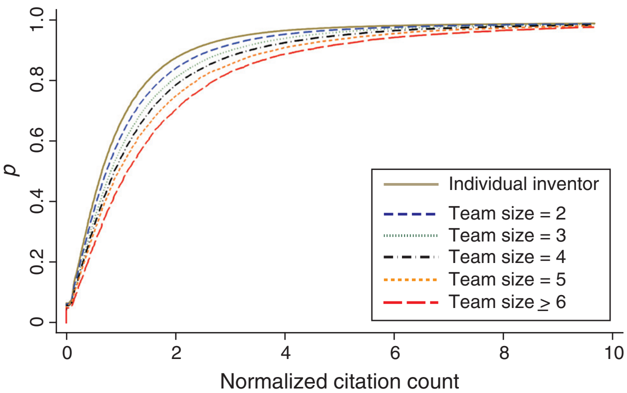
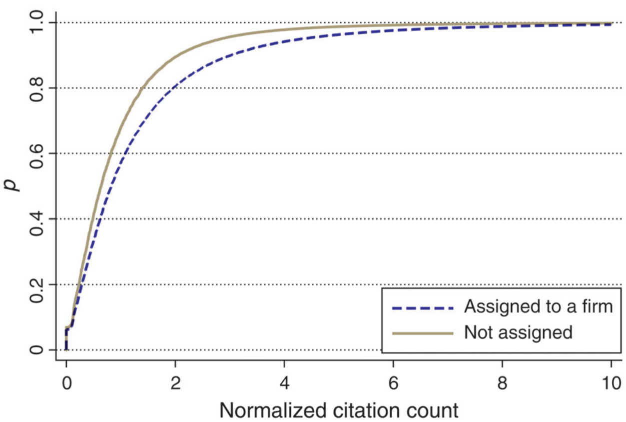


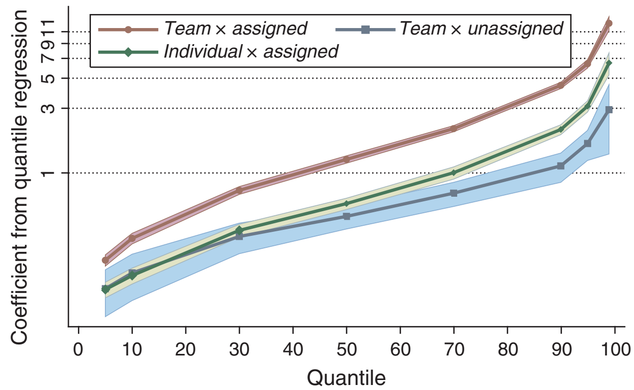
Figure 2(b) **Cumulative Distribution of Observed Impact of Unassigned vs. Assigned Patents**



statistics suggest that team and organization affiliation is *more likely* to be associated with breakthrough (i.e., high-citation) outcomes while simultaneously being *less likely* to be associated with low-value (i.e., low-citation) outcomes. The raw data appear to be most consistent with the theoretical scenario (iv) in Figure 1(b).

Figures 2(a) and 2(b), respectively, illustrate the cumulative distribution of observed outcomes from individuals versus teams (of different sizes) and from unaffiliated inventors versus inventors whose patents are assigned to organizations.[3] Notice that the cumulative distribution plots in both figures *never* intersect. In Figure 2(a), the outcome distribution for teams stochastically dominates the distribution for the lone inventor, with larger teams actually also dominating smaller teams. Likewise, in Figure 2(b), the outcome distribution for patents assigned to organizations stochastically dominates the distribution for

Figure 3 **Relevant Coefficients and 95% Confidence Intervals from Quantile Regression of Citations on Lone Invention**



unassigned patents. Overall, these figures are once more consistent with a view that being a lone inventor decreases the probability of breakthroughs while increasing the probability of particularly bad outcomes. Again, the evidence appears most consistent with scenario (iv) from Figure 1(b). The next section introduces regression models that enable us to examine these issues in more detail and to formally test the hypotheses stated earlier.

4. Regression Methodology

Whereas previous research has primarily focused on effect of collaboration on the *average* outcome, we are interested in the entire distribution of outcomes. In particular, we start by examining drivers of extreme outcomes—inventions that can be considered breakthroughs versus those with little impact. We estimate logistic regression models of the likelihood that a patent's impact falls within one of these two extremes. At the upper tail, breakthrough inventions are defined using an indicator variable *cites_p95* that is set to 1 if and only if a patent is in the top 5% in terms of frequency of future citations received, compared with patents of the same application year and technology class. Analogously, particularly poor outcomes are defined using an indicator variable *citesEQ0* that is set to 1 if and only if a patent receives no citations.[4]

Table 2 summarizes our key variables. Two indicator variables capture two different dimensions of "lone inventor." The first variable, *team*, captures whether a patent came from a single inventor (0) or from a team of two or more inventors (1). The second variable, *assigned*, captures whether the patent originated outside the boundaries of any organization (0)

[3] To enable valid comparisons across different technologies, these figures are drawn using a citation impact measure that has been normalized relative to the average citation impact in the same year-technology class cohort.

[4] All findings reported in the paper are robust to using the top 1% citation impact in defining breakthroughs and also to using achievement of either 0 or 1 future citations in defining particularly poor outcomes.

Table 2 Variable Definitions and Summary Statistics

		Mean	Std. dev.	Min	Max
Dependent variables					
Cites_p95	Patent top 5% in citation impact	0.05	0.22	0	1
CitesEQ0	Patent receives no citations	0.07	0.25	0	1
Explanatory and control variables					
Team	Indicator that is 1 if and only if patent invented by more than one person	0.50	0.50	0	1
Assigned	Indicator that is 1 if and only if patent has an assignee firm	0.76	0.43	0	1
Claims	Number of claims made by the patent	14.66	11.78	1	320
Patent_references	Number of backward citations that the patent makes to other patents	10.86	12.08	0	745
Nonpatent_references	Number of nonpatent references made by the patent	1.92	6.59	0	100
Average_experience	Average number of previous patents for this team's inventors	5.57	13.39	0	347
Joint_experience	Number of past patents invented by the same team	1.90	6.71	0	274
Mediator variables					
Experience_diversity	Number of technology classes any team inventor has patented in before	6.16	9.52	0	234
Network_size	Number of inventors at distance ≤ 2 in the team's collaborative network	11.77	30.12	0	991

or from within an organization (1).[5] An interaction of these two variables tests the implicit hypothesis that a doubly isolated inventor—an individual working outside any organization—is at the most severe disadvantage.

The regression analysis employs several control variables suggested by previous research: *claims* (the scope of the patent as measured by its number of claims), *patent_references* (number of references made to previous patents), *nonpatent_references* (number of references made to public sources outside of patents), *average_experience* (the average number of past patents members of this team have been involved with), and *joint_experience* (the number of past patents from the same team). Because all these variables are highly skewed, we use their logarithmic transformation in the actual analysis.[6] These variables control for the greater resources of teams and the possibility of previously developed collaborative advantage. Technology fixed effects were used to account for systematic differences in citation rate across different technologies. Likewise, year fixed effects were used to account for any systematic differences over time, including those arising from different observed "windows of opportunity" to be cited by future patents until 2004. Finally, to account for the possibility that error terms might be correlated for observations involving the same inventor, we report robust standard errors that are clustered on the identity of the first inventor.

To generalize from examination of the two extremes to the entire distribution of outcomes, we employ a quantile regression approach (Koenker and Bassett 1978; for its first application in the study of creativity, see Girotra et al. 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. This allows an explanatory variable to exhibit different effects for different percentiles.

We test the mediator hypotheses with the procedure suggested by Baron and Kenney (1986). For testing Hypothesis 3, that benefits of collaboration operate through the greater diversity of experience that teams bring to bear, we define *experience_diversity* as the number of distinct technology classes the inventor or inventors have patented in before. This assumes that knowledge of more technical areas offers a greater diversity of ideas available for recombination of criticism. For testing Hypothesis 4, that benefits of collaboration operate through indirect collaborative networks, we construct a measure of the external collaboration networks of the inventor(s): *Network_size* for the focal patent is defined as the number of unique inventors that are at a social distance of not more than two from the focal inventor(s), i.e., are either recent collaborators or collaborators' collaborators for one or more of the inventor(s).

5. Results

A summary of definitions and key statistics for all variables appears in Table 2, and a matrix of correlations among these variables is reported in Table 3. This section details regression analyses.

5.1. Lone Inventors and Extreme Outcomes

The analysis reported in columns (1)–(4) in Table 4 demonstrates that patents generated by inventors with a team and/or organization affiliation are more likely to end up as breakthroughs than those generated by lone inventors. The magnitudes of these

[5] All our findings are robust to using the actual number of inventors behind a patent rather than just an indicator variable (*team*) for this number being 0 (for an individual inventor) or more than 1 (for a team). We report the latter for ease of direct comparison with the other variable (*assigned*, which is also an indicator variable).

[6] In doing so, we first added one to any variables that can take a value of 0. The results are robust to changing the size of the offset or using the original variable.

Table 3 Correlation Matrix Among Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) *Cites_p95*	1.000										
(2) *CitesEQ0*	−0.062	1.000									
(3) *Team*	0.058	−0.011	1.000								
(4) *Assigned*	0.058	−0.012	0.344	1.000							
(5) ln_*claims*	0.089	−0.071	0.100	0.136	1.000						
(6) ln_*patent_references*	0.067	−0.082	0.020	0.002	0.156	1.000					
(7) ln_*nonpatent_references*	0.074	0.011	0.164	0.168	0.132	0.089	1.000				
(8) ln_*average_experience*	0.029	0.036	0.156	0.272	0.072	0.016	0.115	1.000			
(9) ln_*joint_experience*	−0.013	0.038	−0.286	0.010	−0.005	0.016	0.002	0.655	1.000		
(10) ln_*experience_diversity*	0.057	0.012	0.343	0.337	0.101	0.018	0.149	0.873	0.436	1.000	
(11) ln_*network_size*	0.069	0.014	0.414	0.380	0.096	0.006	0.206	0.628	0.070	0.669	1.000

effects are substantial. For example, the estimates in column (3) imply that, keeping other variables at their average value, patents from teams are 28% more likely than patents from individuals to be in the 95th percentile of citations. Similarly, assigned patents are 63% more likely than unassigned patents to be in the 95th percentile of citations.[7] Column (4) examines interaction effects between team and organization affiliation and finds the two to be complements: The citation impact is greatest for patents arising from teams associated with organizations, with such patents being 2.11 times as likely to achieve a breakthrough compared with a lone inventor with neither team nor organization affiliation. There appears to be little evidence that lone inventors are the sources of breakthroughs.[8]

Next, we examine the other extreme of how being a lone inventor affects the likelihood of inventing particularly poor outcomes. As the results report in columns (5)–(8) of Table 4, patents from inventors with team and/or organization affiliation are less likely to receive no citation at all. The magnitude of these effects is again substantial. For example, the estimates in column (7) imply that patents from teams are 9% less likely than patents from individuals to receive no citations. Similarly, assigned patents are 14% less likely than unassigned patents to receive no citations. Column (8) examines interaction effects between working in teams and working in an organization and finds patents arising from teams associated within organizations to be the least likely to fail (they are 22% less likely to have no citations compared with a lone inventor with neither team nor organization affiliation). Overall, lone inventors seem more likely to end up in the left tail of the overall distribution of outcomes.[9]

To summarize, the evidence rejects a view that being a lone inventor increases the probability of achieving both extremely good and extremely bad outcomes. Instead, collaboration (both in terms of team affiliation and organization affiliation) is beneficial at both extremes: It *increases* the probability of breakthroughs while simultaneously *decreasing* the probability of particularly poor outcomes. More generally, the findings demonstrate why an analysis of only working alone versus collaborating achieves greater variance and would be misplaced because it ignores the very real possibility that breakthroughs need not come at the expense of simultaneously increasing the likelihood of poor outcomes.

5.2. Quantile Regression Analysis

Figure 3 plots the estimated coefficients and associated 95% confidence intervals for the first three indicator variables from a quantile regression (Koenker and Bassett 1978).[10] This helps compare three categories of patents—assigned patents from teams, assigned patents from individuals, and unassigned

[7] We have been somewhat conservative in using the number of claims, number of patent references, and number of nonpatent references as control variables. A more aggressive interpretation could be that these three variables are also potential mediators through which working in a team or an organization, or both, shapes the final outcome. Indeed, these variables are positively correlated with likelihood of a patent being a breakthrough, and excluding them further increases the estimated effects of *team* and *assigned*, casting lone inventors in an even poorer light.

[8] This finding is robust to redefining breakthroughs based only on citation counts calculated even after dropping citations an assigned patent receives from future patents originating within the same organization. In other words, the results are not just a manifestation of assigned patents generating significant within-organization citations.

[9] This finding is also robust to dropping citations within the organization as far as the team versus individual inventor distinction is concerned. However, the finding that assigned patents are less likely to end up in the left tail than unassigned patents no longer holds if assignee self-citations are excluded, suggesting that the impact of assigned patents with only a few citations is disproportionately likely to be confined within the organization itself.

[10] To conserve space, the actual table of quantile regression estimates (including those for the control variables) has not been included in the paper, but is available from the authors upon request. Also note that in Figure 3 a logarithmic scale has been used for the y-axis for easy visual comparison of different curves even at the bottom tail of the distribution.

Table 4 Regression Analyses of Extreme Outcomes upon Lone Invention

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	*Cites_p95*	*Cites_p95*	*Cites_p95*	*Cites_p95*	*CitesEQ0*	*CitesEQ0*	*CitesEQ0*	*CitesEQ0*
Regression model:	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic
Team	0.347**		0.257**		−0.125**		−0.0962**	
	(0.019)		(0.019)		(0.017)		(0.017)	
Assigned		0.573**	0.508**			−0.182**	−0.161**	
		(0.025)	(0.026)			(0.016)	(0.017)	
Team × *assigned*				0.780**				−0.262**
				(0.030)				(0.021)
Team × *unassigned*				0.340**				−0.176**
				(0.044)				(0.032)
Individual × *assigned*				0.535**				−0.184**
				(0.031)				(0.019)
ln_*claims*	0.473**	0.462**	0.457**	0.458**	−0.309**	−0.302**	−0.302**	−0.302**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.0088)	(0.0088)	(0.0088)	(0.0088)
ln_*patent_references*	0.243**	0.244**	0.239**	0.239**	−0.315**	−0.314**	−0.314**	−0.314**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.010)	(0.0100)	(0.010)	(0.010)
ln_*nonpatent_references*	0.284**	0.283**	0.276**	0.276**	−0.0998**	−0.0984**	−0.0957**	−0.0959**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
ln_*average_experience*	0.146**	0.161**	0.111**	0.111**	−0.0111	−0.0165	0.00463	0.00300
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
ln_*joint_experience*	−0.135**	−0.212**	−0.119**	−0.121**	0.123**	0.149**	0.115**	0.118**
	(0.020)	(0.019)	(0.020)	(0.020)	(0.016)	(0.013)	(0.016)	(0.016)
Year fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Technology fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Observations	509,840	509,840	509,840	509,840	509,840	509,840	509,840	509,840
χ^2	4,960	5,341	5,318	5,301	12,601	12,637	12,693	12,698
Degrees of freedom	50	50	51	52	50	50	51	52
Log likelihood	−100,915	−100,711	−100,577	−100,575	−113,203	−113,172	−113,150	−113,145

Note. Robust standard errors clustered by the first inventor are in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$.

patents from teams—against unassigned patents from individuals as the omitted (reference) category. The results demonstrate that along the entire distribution, having both team and organization affiliation dominates having only one of the two kinds of affiliations, which in turn dominates being a lone inventor not affiliated with any team or organization. Once more, there is no evidence of lone inventors performing better in any part of the distribution. The effects between adjacent quantiles are significantly different, demonstrating that this reflects more than a simple mean shift. In fact, the difference across the four categories is significantly larger for the higher quantiles, indicating that lone inventors are particularly *disadvantaged* when attempting to achieve breakthroughs.

5.3. Experience Diversity and Network Size as Mediators

We next employ mediation analysis. The first step, as per Baron and Kenny (1986), is to establish that *team* and *assigned* (the explanatory variables) significantly affect *experience_diversity* and *network_size* (the proposed mediators) in the expected direction. As shown in Table 5, both team and organization affiliation are indeed positively associated with greater diversity of experience as well as network size. The second check is to establish that, in regressions not including the explanatory variables *team* and *assigned*, the potential mediators *experience_diversity* and *network_size* are positively associated with the likelihood of breakthroughs and negatively associated with the probability of particularly poor inventions. This was confirmed to be true in an additional analysis not reported here to conserve space.

As the third and most crucial step, we need to check whether the magnitude of the estimated effect of *team* and *assigned* (the explanatory variables) decreases significantly with inclusion of the mediators. This step is shown for the two dependent variables—*cites_p95* (indicator for being among the top 5% in citation impact) and *citesEQ0* (indicator for getting zero citations)—in columns (1)–(5) and columns (6)–(10), respectively, of Table 6. The difference between regression coefficients for either explanatory variable (*team* or *assigned*) between columns (1) and (4) as well as between columns (6) and (9) is statistically significant,

Table 5 **Regressions of Experience Diversity and Network Size as Potential Mediators**

	(1)	(2)	(3)	(4)
Dependent variable:	*Experience_diversity*	*Experience_diversity*	*Network_size*	*Network_size*
Regression model:	Negative binomial	Negative binomial	Negative binomial	Negative binomial
Team	0.580**		0.723**	
	(0.0054)		(0.012)	
Assigned	0.172**		1.331**	
	(0.0071)		(0.024)	
Team × *assigned*		0.733**		2.146**
		(0.0081)		(0.035)
Team × *unassigned*		0.488**		1.068**
		(0.013)		(0.044)
Individual × *assigned*		0.139**		1.468**
		(0.0087)		(0.035)
ln_*claims*	0.0407**	0.0405**	0.0367**	0.0369**
	(0.0027)	(0.0027)	(0.0070)	(0.0071)
ln_*patent_references*	0.00657*	0.00631*	−0.000752	−0.000112
	(0.0026)	(0.0026)	(0.0055)	(0.0055)
ln_*nonpatent_references*	0.0297**	0.0295**	0.0882**	0.0876**
	(0.0024)	(0.0024)	(0.0053)	(0.0053)
ln_*average_experience*	1.064**	1.065**	1.689**	1.688**
	(0.0049)	(0.0049)	(0.0089)	(0.0090)
ln_*joint_experience*	−0.114**	−0.114**	−0.983**	−0.979**
	(0.0057)	(0.0057)	(0.012)	(0.012)
Year fixed effects	Included	Included	Included	Included
Technology fixed effects	Included	Included	Included	Included
Observations	509,840	509,840	509,840	509,840
χ^2	104,975	105,715	83,382	82,695
Degrees of freedom	51	52	51	52
Log likelihood	−1,113,549	−1,113,415	−1,194,324	−1,193,965

Note. Robust standard errors clustered by the first inventor are in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$.

with the implied marginal effect also falling substantially in both cases.[11] We also look for possible interaction effects between experience diversity and network size. Although we find no significant interaction between the two for achieving breakthroughs (column (5)), they are found to be substitutes in avoiding particularly poor outcomes (column (10)).

Strictly speaking, coefficient estimates are not directly comparable across different logistic models, and additional analysis is needed to ensure that the economic magnitude of the mediation effect is substantial and in the expected direction. We found that once the mediator variables are introduced, the increase in probability of a breakthrough falls from 28% to 10% for *team* and from 63% to 51% for *assigned*. Similarly, the associated decrease in the probability of a zero citation outcome falls from 9% to 2% for *team* and from 14% to 10% for *assigned*. Taken together, these results consistently suggest that once the two mediators are included, benefits associated with team or organization affiliation decrease significantly in both cases. In other words, gains associated with team or organization affiliation appear to operate through these mediators to a significant extent.

The mediators do a better job of explaining benefits from team affiliation than from organization affiliation, perhaps not a surprise given that they are both team-level constructs. Interesting differences also exist in the relative importance of experience diversity versus network size at the top versus bottom extreme of the distribution.[12] We find that the ratio of regression coefficients for experience diversity over network size is significantly smaller for the top extreme (column (4)) than for the bottom extreme (column (9)), suggesting that the relative gains from experience diversity (when compared to those of network size) are less important for breakthroughs than for poor outcomes. A similar statement can be made about

[11] Because coefficient estimates from different regression models are not independent (Clogg et al. 1995), we carried out Wald tests to compare coefficients between different models using the seemingly unrelated estimation procedure (available as *suest* in Stata 10).

[12] The analysis reported here derives from a series of tests comparing coefficients across models, implemented using nonlinear hypothesis testing (*testnl* in Stata 10) after a seemingly unrelated estimation procedure (*suest* in Stata 10).

Table 6 Mediation of Lone Invention by Experience Diversity and Network Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	*Cites_p95*	*Cites_p95*	*Cites_p95*	*Cites_p95*	*Cites_p95*	*CitesEQ0*	*CitesEQ0*	*CitesEQ0*	*CitesEQ0*	*CitesEQ0*
Regression model:	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic	Logistic
Team	0.257**	0.137**	0.202**	0.102**	0.101**	−0.0962**	−0.0363*	−0.0848**	−0.0287	−0.0249
	(0.019)	(0.020)	(0.019)	(0.021)	(0.021)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)
Assigned	0.508**	0.487**	0.447**	0.435**	0.429**	−0.161**	−0.150**	−0.146**	−0.138**	−0.109**
	(0.026)	(0.026)	(0.026)	(0.026)	(0.027)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
ln_*experience_diversity*		0.331**		0.286**	0.279**		−0.187**		−0.182**	−0.132**
		(0.019)		(0.019)	(0.020)		(0.016)		(0.016)	(0.017)
ln_*network_size*			0.198**	0.177**	0.180**			−0.0470**	−0.0404**	−0.0604**
			(0.0091)	(0.0091)	(0.0091)			(0.0081)	(0.0081)	(0.0085)
ln_*expdiversity* × *ln_netsize*					−0.00739					0.0452**
					(0.0068)					(0.0052)
ln_*claims*	0.457**	0.452**	0.462**	0.457**	0.456**	−0.302**	−0.298**	−0.303**	−0.299**	−0.295**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.0088)	(0.0088)	(0.0088)	(0.0088)	(0.0088)
ln_*patent_references*	0.239**	0.238**	0.238**	0.238**	0.238**	−0.314**	−0.313**	−0.314**	−0.314**	−0.316**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.010)	(0.0100)	(0.0100)	(0.0099)	(0.0099)
ln_*nonpatent_references*	0.276**	0.270**	0.266**	0.261**	0.261**	−0.0957**	−0.0920**	−0.0920**	−0.0890**	−0.0881**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
ln_*average_experience*	0.111**	−0.197**	−0.129**	−0.369**	−0.365**	0.00463	0.173**	0.0611**	0.217**	0.187**
	(0.012)	(0.023)	(0.015)	(0.024)	(0.024)	(0.011)	(0.019)	(0.014)	(0.021)	(0.021)
ln_*joint_experience*	−0.119**	−0.0616**	0.0488*	0.0831**	0.0786**	0.115**	0.0871**	0.0742**	0.0530**	0.0820**
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)
Year dummies	*Included*	Included	Included	Included	Included	Included	Included	Included	Included	Included
Technology dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Observations	509,840	509,840	509,840	509,840	509,840	509,840	509,840	509,840	509,840	509,840
χ^2	5,318	6,178	5,612	6,465	6,559	12,693	12,652	12,769	12,736	12,795
Degrees of freedom	51	52	52	53	54	51	52	52	53	54
Log likelihood	−100,577	−100,251	−100,133	−99,896	−99,895	−113,150	−113,040	−113,123	−113,021	−112,970

Note. Robust standard errors clustered by the first inventor are in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$.

their relative role in the mediation results for breakthroughs versus poor outcomes. Focusing first on the *team* variable, the extent of the drop in coefficient magnitude for *team* is smaller in going from column (1) to column (2) versus column (3) than in going from column (6) to column (7) versus column (8). The mediating role of experience diversity (relative to that of network size) appears to be less prominent for achieving breakthroughs than for avoiding particularly poor outcomes. An analogous comparison holds for the *assigned* variable as well, though the result in this case is driven more by the mediation role of network size for getting breakthroughs rather than by either mediator having much role in avoiding particularly bad outcomes. These significant differences highlight the importance of examining each tail of the distribution separately. They also call into question the common argument that diversity leads to breakthroughs by enabling wider recombinant search. Instead, the beneficial effect of diversity on invention may occur through an improvement of selection processes.

6. Discussion

Building on a model of evolutionary search with its three classic phases of variation, selection, and retention, we argued that collaboration in the form of team and/or organization affiliation enables more careful and rigorous selection of the best ideas while also increasing the combinatorial opportunities for novelty. Inventors affiliated with teams or organizations, or both, should therefore be less likely to create useless inventions and more likely to create breakthroughs. Analysis of an archival data set of patent data supported these arguments; collaboration appears to trim the lower tail of the distribution of citations to a patent and thicken the number of far-right outliers. Inventors with both team and organization affiliation had higher impact across the entire distribution of outcomes, compared with those that were collaborative on only one of these dimensions, whereas individual inventors working outside the context of any organization were found to have the worst impact across the entire distribution. The diversity of experience and extended collaborative networks of the inventor(s) partially mediated gains from team and/or organization affiliation at either extreme. Intriguingly, the effect of diversity appears to be relatively more important in trimming poor outcomes, whereas the effect of an extended collaboration network appears relatively more important in the creation of outlier breakthroughs.

Although our archival database enables this study by furnishing enough data to model extremely rare events, it also imposes a number of limitations. As is typical of most studies on collaboration that employ archival data, our results are based on a cross-sectional comparison of preexisting (and successful) teams. Compared to laboratory teams, analysis of preexisting and real-world teams provides greater external validity to our results, but at the cost of not having a random assignment of inventors to individual versus collaborative effort in the form of team and/or organization affiliation. Because we do not have a model of how individuals get assigned to specific teams or organizations, we cannot be sure if the same inventor(s) would have behaved according to our predictions in different situations. Selection issues also remain, because publication data by definition only record successful submissions. Similar unobserved influences might also complicate the number of inventors that claim coauthorship of a breakthrough; when a particularly promising invention is applied for, more peripheral inventors may claim to have contributed to the work and thus inflate the explanatory variable.

One specific concern with this cross-sectional approach is that better or well-established inventors may have a greater propensity to engage in collaboration. This could lead to an upward bias in estimated gains from collaboration. We tried to address this issue through three robustness checks. First, to capture the possibility that certain inventors might have a greater affinity for collaboration, we analyzed models that explicitly accounted for past projects for the inventor(s) of a patent being disproportionately collaborative. This did not substantively alter our results. Second, we analyzed a subsample consisting only of inventors with no prior patents and found even stronger advantage to collaboration. Third, we reran the analysis employing panel data models with fixed effects meant to capture time-invariant characteristics of the first inventor. The main findings remained unchanged. Admittedly, even fixed effects models are insufficient for ruling out unobserved individual characteristics that change over time. A natural experiment or an instrumental variable approach would have been ideal to address many of these concerns, but neither was found in our research context.

Another concern, which applies not just to our study but to most existing research on collaboration, is the difficulty of making conclusive statements about the *net* value of collaboration. In other words, would a particular set of inventors be more likely to invent a breakthrough if they collaborated from the start or if they worked alone and then pooled their efforts at the end? (Or following Girotra et al. 2007, at what point in the process should inventors start working together?) 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—assumedly, such communication is instantaneous and almost costless within an individual). This question strikes as particularly important for future work, from both a theoretical and normative perspective.

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 advantage to collaboration 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, because of the rarity of creative outliers), such a lab study would still be unable to provide the long-term and social impact of breakthroughs (as measured by citations) that occur and diffuse in the real world.

At a minimum, we hope this article prompts a rethink of current approaches to researching creativity. Theorists should increase their metaphorical "degrees of freedom"; rather than focusing only on the mean or the symmetric variance, or both, of a distribution, they should begin thinking about the different effects a variable may have on the lower tail, the mean, and the upper tail. Diversity, for example, has been argued and shown to have frustratingly wide and conflicting influences upon creativity. Proponents of diversity have argued and shown that the collaboration of diverse individuals creates opportunities for creative abrasion and technological brokering (Leonard-Barton and Swap 1999, Burt 2004); detractors of diversity have argued and shown that diversity leads to communication and interpersonal problems (Dougherty 1992); those seeking resolution have argued and demonstrated that diversity helps creativity until the point where the cognitive and communication challenges become too great, which should be observable in a nonmonotonic effect (Ahuja and Lampert 2001). These arguments are all reasonable and interesting. The mixed evidence adduced to date, however, should be reconsidered in light of a differential effect of diversity on the tails, keeping in mind that previous results may have been biased by incorrect model specification, outlier data points, and sensitivity to heteroscedasticity. All of these proposed mechanisms for diversity's effect may operate, but they may do so in different regions of the distribution and in surprising ways (such as an improvement in the lower rather than upper tail).

Our results on external collaboration networks also speak to the controversy of how small world networks influence creativity (small world networks are characterized by individually tight clusters of networks that are linked together sparsely; see Watts and Strogatz 1998). Uzzi and Spiro (2005) demonstrate positive correlations between small world structure and creativity in Broadway musicals, and Fleming et al. (2007) demonstrate weak correlations of both signs between small world structure and (1) subsequent patenting in regions and (2) citation and future use of a new creative combination. This conflicting and weak evidence in prior empirical work on small worlds may be due to a failure to look at the tails of the creative distribution. Consistent with our results on the importance of extended networks to the invention of breakthroughs, extended networks may increase the skewness of breakthroughs due to enhanced contagion effects (Bikhchandani et al. 1992). Alternatively, if one thinks of the last evolutionary stage of "retention" as the aggregation of "selection" choices made by each member of the larger community, then large and extended networks might increase the proportion of failures as well. This would not occur from differences in the intrinsic quality of ideas but from the faster recognition and identification of poor versus promising ideas. Dense and extended social networks would enable faster diffusion of tacit and inside information, and this might allow for faster sorting of useless versus promising novelty. Note that this argument differs from an argument that extended networks should increase average impact, because of an increased number of social conduits for diffusion. How large social networks influence the retention and distribution of invention outcomes strikes us as an important topic for future research.

Returning to the paper's specific motivation, the conflicting results for lone inventor variance (Dahlin et al. 2004, Fleming 2007 versus Taylor and Greve 2006, Girotra et al. 2007) may have been driven by an increase in lone inventor failures rather than a symmetric increase in failures and breakthroughs. But the models of variance in these conflicting studies implicitly assumed symmetry of the tails and therefore masked any asymmetric influence of collaboration (though Girotra et al. 2007 estimate a quantile model). Although 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 distribution may provide only limited understanding of the phenomenon, particularly if the outcomes vary greatly.

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) 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 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 nor the huge number of trials required to invent an outlier breakthrough. Other methods, such as econometric analyses of large sample archival data, are better suited to understanding breakthroughs and 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. Because 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 et al. 2007, Kavadias and Sommer 2007) and appropriate econometric techniques (e.g., Koenker and Bassett 1978) into their research.

7. Conclusion

Is the lone inventor more or less likely to invent a breakthrough? Using U.S. patent authorship and citation data, we establish empirically that the lone inventor is less likely to invent a breakthrough. This result holds for inventors who work outside of organizations and for inventors who work without coauthors (those who work alone and outside of an organization are least likely of all to invent a breakthrough). Recent and conflicting research on the topic has focused on how lone invention influences the variance of creative outcomes. This research has assumed symmetric influences of lone invention upon the second moment. We advance arguments for the importance of asymmetric influences and demonstrate that lone invention decreases the chances of a breakthrough and increases the chances of a relatively useless invention. Quantile regression supports these arguments and establishes that the effect is more than a mean shift.

Considering the mechanisms that give rise to these results, we propose that collaboration trims the bottom of the distribution, because of a greater number and variety of critical assessments; isolated inventors are less effective than more social inventors at culling the bad ideas. Collaboration should also increase the number of successful outliers because the diversity of groups enables more novel combinations. Supporting these arguments, greater diversity within a group and extended collaboration networks beyond a group appear to mediate the benefits of collaboration. Surprisingly, it appears that diversity has a relatively and significantly greater effect in trimming poor outcomes than fostering breakthroughs. In contrast, extended social networks also appear to be more beneficial to the invention of breakthroughs than to the trimming of poor outcomes.

The results do not support accounts of the heroic lone inventor, at least during the relatively recent time period covered by our study; the story could be a myth, at least when considering the ultimate success of an invention. Although we agree with Steinbeck that original insight occurs within an individual brain, it also appears that creative processes can benefit greatly from collaboration (though perhaps this is also consistent with his elaboration that, "... the group can build and extend it," Steinbeck 1952, pp. 130–131). The stochastic dominance of collaborative efforts over socially isolated efforts is particularly telling; at every point in the distribution of creative outcomes, collaboration correlates with increased impact. Still, we cannot put the myth entirely to rest. It may be that the lone inventors of the 19th century were truly heroic, but 20th century changes in technology and organizing have increased the advantages of collaboration. It is also possible that the opportunity costs of inventing together may be greater than the costs of simply working alone and then summing the individual efforts. In either case, the larger goal of this paper was to motivate a reexamination of how we study creativity. As illustrated by the significantly different influence of diversity on the top and bottom of the creative distribution, focusing solely on the mean—or assuming symmetric influence of collaboration upon the tails of a distribution—can be methodologically insufficient and substantively misleading. However future researchers think about and study these issues, we hope this work highlights the importance of studying the asymmetric moments of collaborative creativity.

Acknowledgments

The authors thank the Harvard Business School Division of Research and INSEAD for supporting this work; Henrich Greve, Eric Lin, Chris Liu, Matt Marx, Kristen Merrill, Santiago Mingo, and Scott Stern for their comments; and participants in presentations at Stanford University, Harvard University, the University of Toronto, and the Academy of Management and INFORMS annual conferences for their feedback. Departmental editor Christoph Loch, an associate editor, and two anonymous reviewers provided exceptional help in developing the work.

References

Ahuja, G., C. Lampert. 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management J.* **22**(6–7) 521–543.

Albert, M. B., F. Narin, D. Avery, P. McAllister. 1991. Direct validation of citation counts as indicators of industrially important patents. *Res. Policy* **20** 251–259.

Baron, R., D. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personality Soc. Psych.* **51**(6) 1173–1182.

Basalla, G. 1988. *The Evolution of Technology*. Cambridge University Press, Cambridge, MA.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* **100**(5) 992–1026.

Brefka, P. 2006. Personal interview, July 10, Boston.

Burt, R. 2004. Structural holes and good ideas. *Amer. J. Sociol.* **110**(2) 349–99.

Campbell, D. 1960. Blind variation and selective retention in creative thought as in other knowledge processes. *Psych. Rev.* **67** 380–400.

Clogg, C., E. Petkova, A. Haritou. 1995. Statistical methods for comparing regression coefficients between models. *Amer. J. Sociol.* **100** 1261–1293.

Csikszentmihalyi, M. 1999. *Creativity*. Harper Collins, New York.

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

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

Diehl, M., W. Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *J. Personality Soc. Psych.* **53** 497–509.

Dougherty, D. 1992. Interpretive barriers to successful product innovation in large firms. *Organ. Sci.* **3**(2) 179–202.

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

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

Fleming, L. 2007. Breakthroughs and the "long tail" of innovation. *Sloan Management Rev.* **49**(1) 69–74.

Fleming, L., C. King, A. Juda. 2007. Small worlds and innovation. *Organ. Sci.* **18**(6) 938–954.

Gilfillan, S. C. 1935. *The Sociology of Invention*. Follett Publishing, Chicago. (Reprinted 1970. MIT Press, Cambridge, MA.)

Girotra, K., C. Terwiesch, K. Ulrich. 2007. Idea generation and the quality of the best idea. Working paper, INSEAD, Fontainebleau, France.

Grant, R. M. 1996. Toward a knowledge-base theory of the firm. *Strategic Management J.* **17** 109–122.

Hall, B., A. Jaffe, M. Trajtenberg. 2005. Market value and patent citations. *RAND J. Econom.* **1** 16–38.

Harhoff, D., F. Narin, F. Scherer, K. Vopel. 1999. Citation frequency and the value of patented inventions. *Rev. Econom. Statist.* **81**(3) 511–515.

Hooker, C., J. Nakamura, M. Csikszentmihalyi. 2003. The group as mentor: Social capital and the systems model of creativity. P. Paulus, B. Nijstad, eds. *Group Creativity: Innovation Through Collaboration*. Oxford University Press, New York, 225–244.

Hughes, T. 2004. *American Genesis: A Century of Invention and Technological Enthusiasm 1870–1970*. Chicago University Press, Chicago.

Jaffe, A., M. Trajtenberg. 2002. *Patents, Citations and Innovations: A Window on the Knowledge Economy.* MIT Press, Cambridge, MA.

Janis, I. 1972. *Victims of Groupthink*. Houghton Mifflin Company, Boston.

Kavadias, S., S. Sommer. 2007. The effects of problem structure and team expertise on brainstorming effectiveness. Working paper,

Kenney, S. 2006. Personal interview, August 13, Cambridge, MA.

King, G. 1989. Variance specification in event count models: From restrictive assumptions to a generalized estimator. *Amer. J. Political Sci.* **33**(3) 762–84.

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

Kogut, B., U. Zander. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* **3**(3) 383–397.

Leonard-Barton, D., W. Swap. 1999. *When Sparks Fly: Igniting Creativity in Groups.* Harvard Business Press, Boston.

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

March, J. G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

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

Mokyr, J. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press, New York.

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

Nelson, R., S. Winter. 1982. *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA.

Office of Technology Assessment and Forecast. 1976. U.S. Department of Commerce, Patent and Trademark Office, Sixth Report. U.S. Government Printing Office, Washington, DC.

Paulus, P., V. Brown. 2003. Enhancing ideational creativity in groups: Lessons from research and brainstorming. P. Paulus, B. Nijstad, eds. *Group Creativity: Innovation Through Collaboration*. Oxford University Press, New York, 110–136.

Paulus, P., B. Nijstad, eds. 2003. Group creativity: An introduction. *Group Creativity: Innovation Through Collaboration*. Oxford University Press, New York, 3–12.

Perry-Smith, J. E., C. E. Shalley. 2003. The social side of creativity: A static and dynamic social network perspective. *Acad. Management Rev.* **28** 89–106.

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

Romer, P. 1993. Ideas and things: The concept of production is being re-tooled. *The Economist* **328**(7828) 70–72.

Runco, M. 1999. Creativity need not be social. A. Montuori, R. Purser, eds. *Social Creativity*, Vol. 1. Hampton Press, Creskill, NJ, 237–264.

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

Schumpeter, J. A. 1934. *The Theory of Economic Development.* Harvard University Press, Cambridge, MA.

Simonton, D. 1985. Quality, quantity, and age: The careers of 10 distinguished psychologists. *Internat. J. Aging Human Development* **21** 241–254.

Simonton, D. K. 1999. *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press, New York.

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

Singh, J. 2007. Asymmetry of knowledge spillovers between MNCs and host country firms. *J. Internat. Bus. Stud.* **38**(5) 764–786.

Singh, J. 2008. Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Res. Policy* **37**(1) 77–96.

Smith, A. 1766. *Lectures on Jurisprudence*. (Reprinted 1982. *Lectures on Jurisprudence* (*Glasgow Edition of the Works and Correspondence of Adam Smith, Vol. 5*). Liberty Fund, Indianapolis.)

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

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

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

Trajtenberg, M. 1990. A penny for your quotes: Patent citations and the value of innovations. *RAND J. Econom.* **21** 172–187.

Trajtenberg, M., G. Shiff, R. Melamed. 2006. The "names game": Harnessing inventors' patent data for economic research. NBER Working Paper 12479, National Bureau of Economic, Research, Cambridge, MA.

Uzzi, B., J. Spiro. 2005. Collaboration and creativity: The small world problem. *Amer. J. Sociol.* **111** 447–504.

Watts, D., S. Strogatz. 1998. Collective dynamics of small-world networks. *Nature* **393** 440–442.

Weitzman, M. 1998. Recombinant growth. *Quart. J. Econom.* **113** 331–360.

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