Selection of Early-Stage Ventures: Social Influence and its impact on Committee Performance[1]

Daniel Fehder
USC Marshall School

Fiona Murray
MIT Sloan and NBER

Abstract: We study the selection of early-stage startups in a top accelerator. The final selection of startups in our setting is made by a rank ordering of startups based on scores produced by committees of judges. We use the random allocation of judges to committees and committees to startups to explore the impact of committee composition on the individual decision-making of each judge. Specifically, we seek to measure whether the perceived expertise of some judges based on professional background or status characteristics exerts differential social influence on other judges. Leveraging the availability of each individual judge's score for each startup, we explore the extent of social influence through a series of regressions estimating the differential peer-effects of judges in and out of a number of group categories. We find that judges with some professional backgrounds, such as early-stage investors, exert a marked social influence on other judges while other backgrounds, such as lawyers, have the opposite effect. We show that female judges have substantially less social influence than male judges, but that high quality judges do not have a larger social influence on their committees than lower quality judges. These patterns of social influence have performance implications. When committee's show signs of disregarding the opinions of committee members with lower social influence, the predictive performance of the committee's startup evaluation is lower.

PRELIMINARY AND INCOMPLETE

PLEASE DO NOT CIRCULATE

[1] * This paper was made possible through the generous support of the Kauffman Foundation. We gratefully acknowledge the help of many in sharpening this paper including Abhishek Nagaraj, Joe Raffiee, Dan Rock, Josh Krieger, and Yael Hochberg. Many thanks as well to Simone Abegunrin and Caroline Fehder for helping to prepare the data. All Mistakes are our own.

Introduction

From corporate and venture capital boardrooms to NIH study sections, a key determinant of an organization's performance is choosing which projects to support. One of the most commonly observed mechanisms for selecting projects is through committees especially in settings where the future performance of proposed projects is highly uncertain such as whether new startups, product lines, or research programs. The prevalence of committee decision making should not be surprising as a broad set of research suggests that more heads are better than one across a wide range of tasks, yet we still know little about how committee composition and internal dynamics impact the performance of evaluation. On the one hand, a broad range of views can improve the performance of the group (Wooley et al., 2010; Mannes, Larrick and Soll, 2012). On the other hand, the outsized influence of a few members of the committee might damage the functioning of the committee in synthesizing potentially disparate points of view.

We use the lens of social influence to explore the extent to which the perceived expertise and social status of committee members shapes the evaluation of early-stage startups, a high uncertainty domain where high performance represents a critical capability across a broad range of organizations. The successful aggregation of a diverse set of opinions seems to drive group performance broadly (Hong and Page, 2004) and in the evaluation process more specifically (Criscuolo et al., 2017). Despite the promise of group diversity, differences in group members introduces the possibility that the opinions of some members might be viewed as more salient based on their perceived status or authority (Cialdini and Goldstein, 2004). Indeed, these types of social cues are particularly important in decision-making in high-uncertainty environments (Rao, Greve and Davis, 2001; Botelho and Abraham, 2017). Can the social influence of high-status individuals (partially) arrest the benefits of group-level diversity in the performance of evaluation committees?

We seek to open the black-box of interpersonal dynamics in evaluative committees because much of the empirical and theoretical work to date has focused more on individual-level decision processes. Substantial work has uncovered preferences and biases that stem from an individual's previous experience, economic interests, and personal characteristics on their evaluation of new projects (Li, 2017; Boudreau et al., 2016). While more recent work has explored the extent to which the opinions of others influences the decision-making processes of evaluators (Becker et al., 2017; Teplitskiy et al., 2019), this empirical work has attempted to isolate the information effects from the potential impact of social influence by restricting the interaction of individual evaluators and not allowing face-to-face engagement. While these experimental studies provide a key baseline for the impact of information about the opinions of other evaluators, it is important to evaluate the functioning of evaluation committees in their natural habitat: the boardrooms where face-to-face interactions predominate.

To do so, we exploit data from the selection process from MassChallenge, a startup accelerator, where committees choose 128 startups each year to enter the program but where we observe each individual judge's scoring behavior. Startup Accelerators are an increasingly important organizational form within entrepreneurship affecting how venture capitalists and corporations interact with startups (Cohen et al., 2019). We explore the evaluation of 1,173 projects by 371 judges that are randomly assigned into committees which are then additionally randomly assigned a set of startups to evaluate using a standardized score sheet. The MassChallenge judges are drawn from a number of groups that exist in the entrepreneurial ecosystems including venture capitalists, experienced entrepreneurs, corporate innovation managers with an interest in startups, and lawyers with a startup clientele. We use this setting to explore the extent of social influence across and within these different professional groups as well as the impact of a key status characteristic, gender. We contrast the potential social influence from professional background and gender with the influence exhibited by high quality judges, an attribute that is unobserved by the committee members but can be derived by the scoring behavior of the judges in earlier rounds of startup selection. When committees work well to process varying points of view, the influence of high-quality judges should exceed that of judges with high-status professional backgrounds or personal characteristics.

To measure the impact of social influence between members in a committee, we propose a novel application of peer effects. To do so, we explore how similar each individual judge's scoring behavior is to other members of the committee through a series of peer effects regressions. In so doing, we characterize social influence as the extent of convergence between an individual's score and that of their socially proximate committee members or those perceived as high status. Next, we explore whether the extent of convergence and divergence of opinion across different groups within a committee predict higher or lower performance of a committee's final scores in predicting a startup's future performance.

Our study provides a number of findings. First, we show that at least in our early-stage venture selection context, there exist very large performance differences across the individual judges in their ability to evaluate the quality of the startups. These quality differences are not tied to the professional backgrounds of the judges. Second, despite the lack of connection between professional background and the individual quality of a judge's judgements, we find substantial differences in the degree to which judges of different backgrounds seem to exert peer influence on the evaluation of other judges. We find that scores of investors have the strongest relationship to the scores of other judges in their committee (and lawyers have the least). In addition, we show that judges are more significantly influenced by judges with whom they share professional background. Lastly, we show that the average judge is not influenced more by high quality judges than they are by lower quality judges. High quality judges, however, are more influenced by other high-quality judges in the committee round.

Lastly, we find that these differences in social influence across different groups have performance implications. When there are larger divergences between judges within a committee based on gender, the recommendations of that committee are less likely to predict the future performance of the startups they are evaluating. In contrast, the committee's recommendations are likely to be more predictive when there is less agreement between committee members that are venture capitalists (the high-status profession in this setting) and non-venture capitalists.

Taken together our results suggest that high quality judges are hard to identify based on their individual professional characteristics. Rather, their quality may be discerned over time suggesting that maintaining careful judging records might be a valuable exercise. Second, it seems that there is also value in committee-based decision making: Even though there are clearly potential channels for bias, the presence of high-quality judges on committees improves the performance of project selection (as compared to individual selection) through two channels. First is the ability of an individual judge to pick high quality startups when given an application, fits nicely with the signal extraction model that underlies the majority of work that has explored project selection. Second is the ability of high-quality judges to be persuaded by the right people (i.e. other high-quality judges) in committee setting. We believe that our paper provides the first empirical evidence that the performance of a committee in selecting high quality projects is not only driven by the individual performance level of the judges but also by their ability to influence each other's judgements. High performance judges not only show good taste in their individual selection of projects but also good taste in how they are persuaded by the signals of others.

The Role of Social Influence in Economic Valuation

If a committee is more than just a forum to collect individual opinions and judges engage in persuasion and influence, our theoretical account of project selections needs to be enriched to account for the differences in evaluation regime – with the potential for persuasion in some evaluation regimes but not others. It must account for process ranging from individual judges selecting projects based on gut feel (Huang and Pearce, 2015) to multi-staged selection committees with formal scoring rules (Kerr, Lerner and Schoar, 2011) and explore the performance implications of influence and persuasion amongst judges. This is of growing importance given the expanding scope of project selection particular in the start-up venture world, and the role of selection in programs such as accelerators, business plan competitions, shark-tank like pitching contests etc.

Effective selection of new, highly uncertain and innovative projects from among a large number of possible options can be conceptualized as decision-making under uncertainty a topic that is not new to the literature in management science. Indeed, the Carnegie School's initial program called careful attention to the use and communication of information by agents in decision making processes with limited attention and processing power in organizations (Cyert and March, 1963; Simon, 1957). Since the initiation of the study of organizations from a behavioral perspective, more attention has been paid to the ways in which firms explore new opportunities than the ways in which firms select amongst them (Knudsen and Leventhal, 2005), but a growing literature has studied the selection of projects within organizations (Csaszar and Eggers, 2013: Criscuolo et al., 2017) and in other institutions such as scientific review and crowdfunding (Boudreau et al., 2016; Li, 2017; Mollick and Nanda, 2015). Nearly all of these studies share a common characteristic that the performance of the group selection mechanism emerges from the quality of match between the individual judges and the project in terms of the information and expertise each judge brings to their evaluation. Thus, the literature has focused on the advantages and potential biases that emerge from this match between evaluator and evaluated. Within this framework, an optimal judging panel is composed of the right individuals, but the aggregation of their information takes place through a static aggregation of evaluations which can also be tuned to fit the right environmental conditions (Csaszar and Eggers, 2013).

While these studies have provided strong evidence that individual committee members, and through them the panel, react strongly to certain characteristics of a project idea depending on the judge's characteristics. Judges react negatively to novelty to varying degrees depending on the distance of the idea to the expertise of the focal judge in some metric of idea space (Boudreau et al., 2016) and this negative relationship seems to expand with the homophily of the committee (Criscuolo et al., 2017). The observed intensification of individual preferences against novelty when interacting in a group suggests that group-level processes might provide more than simply the aggregation of individual preferences and beliefs about the value of a project.

The role of discussion on group-level performance in judgement tasks is an old topic for which there is surprisingly little convergence. On the one hand, there is evidence that the sharing of perspectives can have a negative impact on group-level judgement (Heath and Gonzalez, 1995) while others suggest that group discussion can improve the judgement beyond averaging of individuals (Schulze et al., 2012). The impact of discussion seems to be highly dependent on the nature of the task, where discussion seems to improve the judgement in group-tasks that exhibit higher degrees of uncertainty (Minson et al., 2017). The role of consensus is at the heart of these varying findings. For some tasks in certain settings, the social pressures to reach a quick consensus might swamp the additional benefit of discussion. In other settings, under high uncertainty and cognitive complexity, a well-informed discussion might serve to lift the informational and cognitive constraints facing each individual evaluator.

Empirical Setting: The MassChallenge Accelerator

The effective study of evaluation of early-stage start-up projects confronts significant data challenges. First, it is necessary to have a large number of projects being evaluated over a relatively short period of time. Second, an analytic would like to have long-term performance data that may be tracked so that actual project outcomes at a later date (as a proxy for quality at the time of selection) can be captured. Third it is essential to have a pool of judges whose professional and social characteristics are known. Fourth, judging scores need to be collected and diligently maintained to allow for analysis of at the time decisions – preferably with a scoring scale rather than a binary decision. And finally, it is useful to have judgements being made on the same or similar sets of projects under different selection regimes. Thus, an ideal setting to explore the ways in which judges perform under different evaluation regimes, is one in which an observer would independently vary both judges to committees and committees to evaluation regime.

One particular context where such a wealth of observable information about a large number of projects selected through a multi-judge process may be found is in startup accelerators. In their most common form, such accelerators solicit applications from a wide array of early-stage firms (typically in accordance with their specific application criteria), and then narrow down through extensive judging to a cohort of firms who will be admitted into the program. With their large numbers of applications and judges, their frequent receipt of public and philanthropic funding (or of private investment), and the tech-oriented nature of their leaders, theses accelerators are an idea research site for those interested in study judging practices.

In our setting, as described below, each early-stage startup is randomly allocated to two sets of evaluator committees (whose constituents are randomly allocated) and evaluated in each of these two regimes.[2] Moreover, each judge evaluates multiple startups in both the individual and committee evaluation regimes. Most importantly, in our study, the matching between judge, startup, and evaluation regime is randomized explicitly. Our research design therefore explores the variation in evaluation outcomes for a fixed set of judges and a fixed set of early-stage ventures that are evaluated through two distinct evaluation regimes: individual-based and committee-based evaluation. Exploiting this random allocation of judges and startups, we seek to identify the impact of committee composition on the performance of individual judges. Specifically, we ask if the social influence exerted by some judges but not other has a measurable impact on the process of evaluation of these startups in the committee and whether this impacts the performance of the committee in its capability to deliver high quality recommendations.

In this study, we utilize observational data from one large and well-established entrepreneurship program – MassChallenge - across multiple years. MassChallenge is a startup accelerator founded in Boston. (While it has now expanded to other regions, our data only include

[2] The individual round of judging is the first round of judging. The committee round is the second round and takes place a few weeks after the first round. For this study, we focus on the startups that made it into the second round of judging and thus are evaluated in both regimes.

start-up ventures that applied to MassChallenge Boston). MassChallenge receives yearly thousands of applications from around the world from early-stage ventures that wish to enter their four-month residential program in Boston (and now elsewhere). During the program, startups receive substantial mentoring and education[3], but this paper focuses specifically on the evaluation of these firms prior to their admission (or decline of admission) from the program. For this paper, we will focus on the last stage of evaluation where the startups are scored by judges in a committee setting.

When a judge agrees to participate in selecting candidates for the MassChallenge program, they agree to review a number of startup applications through two rounds. In the first round (individual-based regime), each judge individually receives a number of applications through MassChallenge's online platform. Each judge is asked to score the startup based on this written application. They do not know which other judges are evaluating this each startup in the first round nor do they have information on the evaluations of that startup by other judges. MassChallenge then aggregates the scores of each individual judge through a simple average. Startups in the top 250-300 in this round are then passed on to the second round. In the second round (committee-based regime) the startups are given five minutes to give an in-person (or Skype-based) pitch to the judges and then have a five-minute Q&A period. After these ten minutes with the startup's founding team, the committee are given five minutes to commune and then each judge scores the startup on their own.

The authors have had the opportunity to observe the judging process across multiple committees and have collected substantial qualitative observations of inter-judge persuasion and influence which drove the empirical approach of this paper. Within the committee round of judging there are a number of ways in which the judges influence each other's judgements about the quality of the startups that they are tasked with evaluating. First, they each have the ability to ask questions in the Q&A round as the rules of order in the MC judging process state that each judge has the ability to ask a question before a judge can ask a second question. The questions asked and a judge's interpretation of the answers form much of the basis of the discussion that happens after the startups are asked to leave the room. Frequently, a judge's contributions to the group discussion proceed by connecting the question asked of the startup to their particular expertise in

[3] For more information on the impact of the MassChallenge program itself, see Fehder (2015).

terms of professional background or industry knowledge as the judge provides context for why the believe the startup's answer either did or did not satisfy them. Thus, the social identity of each of the committee members most frequently comes into play during the group discussions that arise during the committee meetings and committee members anchor their comments in their beliefs in their authority and credibility based on their work experience.

Analytical Approach

By following the evaluation behavior of a fixed set of judges that are randomly allocated to committees that evaluate a fixed set of business opportunities, we seek to understand the extent to which social influence shapes the individual decision-making of judges and how these potential social influences shape performance differences across different committee compositions.

In our first set of core regressions, we quantify social influence through a novel application of peer effects regressions using leave-out means of different groups to explore the differential impacts of judges with certain characteristics. We build upon the application of the exploration of peer effects on choice behavior (Angrist 2011) and especially the impact of differential peer groups Carrell, Sacerdote and West, 2013). Like many recent examples of peer effects regressions, we use random assignment to break the automatic link between committee composition and observed similarity in the score choice of committee members that might arise because they have similar tastes and outlooks. Notably, our peer effects regressions are different than their typical use because the outcome of interest, the score chosen by the focal judge, and the key shifters of the judge's choice that we explore in these regressions, the scoring behavior of peers in the committee, are driven by shared observation of the quality of the startup which are only partially observed by the econometrician. Indeed, if we were only exploring the relationship between one judge's score and the leave-out mean of the rest of the committee, we would expect to find a statistically significant relationship even without the committee members speaking to one another. In our exploration of differences between in and out group members across a number of social categories, however, the automatic statistical relationship between shared observation and peer effects breaks down. We explore this point empirically in Appendix A and theoretically in Appendix B.

In our core regression in this paper, therefore, we focus on the relationship between the different leave out means judges in and outside of key group categories (e.g. early stage investors) and the score chosen by a judge as a proxy for the degree to which an individual judge places weight on the opinion different types of judges and estimate the following:

$$S_{ij} = \alpha_j + \beta_1 \overline{X}_{ic(-j)} + \beta_2 \overline{X}_{i(-c)(-j)} + \varepsilon_{ij} \quad (1)$$

Here S_{ij} is the score chosen by judge j for startup i, and α_j is a judge fixed effect. After accounting for the average scoring behavior of each judge, we wish to understand the impact of judges of type c on the scoring behavior of judge j. To do so, we use the leave-out mean calculated for the startup in each round for judges with a certain characteristic (investor background, former entrepreneur, etc), $\overline{X}_{ic(-j)}$, and also calculate the leave-out mean for judges without that characteristic, $\overline{X}_{i(-c)(-j)}$.

Equation 1 measures the average impact of different types of judges on the judging behavior of other committee members across all of the startups and all judges in our sample but it does not account for the possibilities of individuals judges being more or less persuaded by judges with whom they share similarities (i.e. homophily). To explore this possibility, we estimate the following:

$$S_{icj} = \alpha_j + \beta_1 \overline{X}_{ic(-j)} + \beta_2 \overline{X}_{i(-c)(-j)} + \beta_3 \mathbb{I}_c \times \overline{X}_{ic(-j)} + \beta_4 \mathbb{I}_c \times \overline{X}_{i(-c)(-j)} + \varepsilon_{ij} \quad (2)$$

Equation 2 builds upon equation 1 and adds two more terms to the estimating equation which interact the group characteristic leave out means with and indicator function for whether or not judge j shares that characteristic c, $\mathbb{I}_c$. By interacting the group-characteristic leave-out means with this indicator, we achieve flexibly estimated peer persuasion effects for each of the potential relationships in the data with respect to the characteristic in question: the impact of other judges with characteristic c on judge j when judge j shares that characteristic (homophily), the impact of the judges without that characteristic on judge j with the characteristic (non-homophily), and the same two possibilities (homophily and non-homophily) for the judges without characteristic c.

Lastly, we wish to explore the extent to which apparent differences in the degree of social influence observed in a committee make a difference for the ability of that committee to effectively perform their role in evaluating early-stage startups. In order to explore the impact of social influence on the performance of committees, we create a new variable at the committee X startup evaluation level which we call Committee divergence. First, we calculate the group-level means for the scores provided by judges either in or out of the group (e.g. investors vs. non-investors) for each committee for each of the evaluations they make. We then take the absolute value of the difference between in and out-group means; we call this committee divergence. We then classify each committee's final score reported to MassChallenge as having a high or low level of disagreement between groups based on whether that committee's final score for that startup was above or below the mean level of divergence for that group category.

With each committee's decision classified as either high or low divergence across each group category of interest, we then measure whether there is a difference in the performance of committee decisions that are high or low divergence for each of these group categories. To do so, we build upon the work of Li (2017) to estimate the following models:

$$\mathrm{S_{icj}} = \alpha_\mathrm{c} + \beta_j Q_{ij} + X_i' \beta_i + \varepsilon_\mathrm{icj} \quad (3)$$

In these regressions, Q_{ij} is the ex-post revealed quality of the startup (logged funding dollars two years after application minus the average treatment effect for admitted startups). The parameter β_j represents an estimate of the value add of committees with below and above average levels of opinion divergence. In addition, we control for static differences in a committee's decision making across all of the startups they evaluate using a committee-level fixed effect, α_c, and also control for vector of factors which are systematically related to the evaluation of a startup's quality, X_i', such as the characteristics of founders including prior entrepreneurship experience, elite education and other factors.

Results

We try to test the hypothesis that a judge will perform better when they have high quality committee members providing them with quality insights. To test this hypothesis, we create a measure of judge quality (which we call judge value added) using a method developed in Li (2017). Using data at the judge X startup level, we estimated the model $Score_{ij} = \beta_j Q_{ij} + X_i + \epsilon_{ij}$ where Q_{ij} is the ex-post revealed quality of the startup (logged dollars two years after application minus the average treatment effect for admitted startups). The parameter β_j represents an estimate of the value add of each judge across the rounds. Figure 3 provides a kernel density plot of the distribution of judge value added in our sample. It shows that most judges have a value added centered around zero while there is a skewed long tail of judges in the right-hand side of the distribution demonstrating that there are judges with value added capacity that far exceeds the modal judge.

To explore whether an individual judge was better able to discern the quality of a startup when they are in good committees, we looked at the average value added for the other judges on the committee in the committee round and the average value added for the virtual committee formed by the other individuals that evaluated a startup in the individual round of evaluation. We then created the variable *High-Quality Committee* if the average value added of the other committee members was in the top quartile (i.e. the average of other committee members was substantially positive). Model 4 of Table 3 explores the relationship between having high quality peers and the performance of an individual judge's scoring behavior. We find that the serving on a high-quality committee in the committee improves the relationship between a judge's score choice and the quality of the startup by a statistically significant 0.4. This means that a one percent improvement in the underlying quality of the startup translates to a 5% of a standard deviation increase in the score a judge chooses to rate that startup. These results suggest that having high quality peers on your committee substantially improves the performance of an individual judge in their capacity to make recommendations.

But who are these high-quality judges? Are there systematic relationships between a judge's background and their ability to assess the quality of startups? To answer this question, we explore to what extent professional background influences the judge's value added. We code each judge for having had work experience in either directly investing in startups, having started a new firm or having provided legal services (these plus general corporate experience are the four backgrounds judges have). In Model 1, we perform a simple OLS regression to explore any statistical relationship between entrepreneurial background and the judge's value-added score and find no statistical relationship. In Model 2, we perform the same analysis but for judges that have experience making investments in early-stage companies. We find no support for a statistical relationship between judges with active investing experience and judge value add. Lastly, we explore whether there is a relationship between working as a lawyer and the ability to evaluate new firms. Similar to the first three regressions, we see no statistical relationship between working experience as a lawyer and the computed judge value add. While we might not be surprised that lawyers are not overrepresented in the higher portions of the distribution of judge value added, it is somewhat intriguing that direct investing relationship is not more strongly correlated with higher judge value added. On the other hand, other work has suggested that early-stage investors are not reliably able to distinguish between good ideas and great ideas in terms of their investing behavior

(Kerr et al., 2014). The results of Table 4 suggest that good judges come from all of these professional backgrounds and that even in the ranks of high-quality judges we observe people with all three of these backgrounds. Thus, it is not immediately apparent who will be a good judge just by checking their LinkedIn page.

Even if professional background does not seem to be systematically linked to judge quality, it is quite reasonable to believe that other judges have priors about the quality of a judge based on their professional background. For instance, they may believe that judges with experience investing in early-stage companies are more likely to have greater ability to discern quality. To what extent do these priors about the value of professional backgrounds of peers on the committee shape the degree to which they influence an individual judge's scoring behavior? To what extent are judges instead influenced instead by the voices of high-quality judges regardless of their background? We explore these questions in Table 5 where we estimate a series of regression models influenced by the peer effects literature. Model 1 provides the impact of the leave out mean on the score of the individual judge in the paper round. This regression is meant to address a key concern in any peer-effects regression where group level leave-out means are used to predict individual behavior: to what extent is an observed relationship evidence of peer effects (here through the persuasion channel) or just a feature of their similarity by virtue of being in the same group (Angrist, 2014). A point increase in the average score of the rest of the committee is associated with a near 0.2 increase in a judge's score in the paper round where they have no contact with the rest of the committee. If judges had similar opinions on average, we would expect this relationship to be stronger even though the individual judges never speak with one another (or even know the identities of the other judges in their virtual committee). Model 2 contrasts that result with the impact of the leave-out mean in the committee round where a one-point increase in the average score of the rest of the committee leads to a 0.6 point increase in a judge's score (around 4.5% of a standard deviation in score). This suggests that there is potentially persuasion and consensus-building happening relative to the scoring behavior in the first round.

Models 3- 10 decompose the result in Model 2 by looking at the differential impact of judges with different professional backgrounds on the focal judge's scoring behavior. We do this by computing separate leave out means for members of the committee with and without the specific background being explored in the regression and exploring the impact of these different leave-out means on the scoring behavior of each individual judge. MassChallenge judges are notified of the identities of the other members of their committee in advance and use the committee round as an opportunity to network. Thus, we can expect that on average each judge has checked the background of the other judges in their committee round. Similar to the first two models, Model 3 and Model 4 estimate regressions that explores the differential influence of judges with investing background across the two different evaluation regimes. Model 3 shows a positive contribution of the scoring of both investor and non-investors on the score choice of individual judges again demonstrating correlated signals across judges, but we do not see a significant difference between the estimated influence of investors or non-investors evaluating the same company. This lack of statistically significant difference is what we expected to observe in the data, but it makes the findings in Model 4 all the more significant. Model 4 shows that judges with investor backgrounds seem to be given more weight in the opinion formation of other judges. A point increase in the leave-out mean of investors is associated with a 0.34 point increase in the focal judges score whereas the leave-out mean of the judges without investing experience increase the average score of the focal judge by 0.25 points. This difference is significant at the 0.001 level. In contrast, Models 5 and 6 estimate the impact of the average scoring behavior of entrepreneurs and non-entrepreneurs in a committee on the scoring behavior of a judge. Whether in the virtual committee of the individual round or in the committee round, judges do not seem to be more influenced by committee members with an entrepreneurial background.

Interestingly, we find the opposite relationship between lawyers on a committee as we found with investors. In Model 7, we find that the opinions of lawyers are less similar to the average judge in the individual round. In Model 8, we show that lawyers on a committee have a statistically significant smaller effect on the scoring behavior of other judges compared to non-lawyers. Non-lawyers have more than twice the impact on the scoring behavior of other judges. In Models 9 and 10, we perform the same analysis for high quality judges (judges in the top quartile of judge value added) and find that there is not a statistically significant difference in the extent to which these judges influence the scoring behavior or the other judges. These results show that there is evidence of peer influence on scoring behavior in the committee round that greatly exceeds the shared viewpoints that judges have when they read the applications and score them separately. Furthermore, the degree of peer background seems to vary with the professional backgrounds of the judges where investors influence individual judges more than average while lawyers influence other judges less than average. Intriguingly, high quality judges do not seem to be accorded any more influence than other judges, perhaps because their identity as high-quality judges is not readily apparent to them or the other committee members.

Table 5 explored the average impact of different types of committee members on the scoring behavior of individual judges, but there is reason to believe that judges will be more persuaded by the opinions of judges with whom they share professional background. We explore this possibility in Table 6 where we build off of the regressions in Table 5 but now isolate the differential impact of these different professional backgrounds when they are shared or not shared by the focal judge. Models 1, 3, 5, and 7 explore this relationship in the paper round where any observed relationships would be due to shared points of view amongst similar judges instead of arising from peer persuasion. Across all of these models, we observe that the shared points of view of judges with the same professional background are not statistically significantly different from each other. In Model 3 for instance, the interaction term between a judge with an entrepreneurial professional background and the leave-out mean for other judges with entrepreneurial backgrounds is positive and statistically significant but there is no statistically significant difference between the relationship between leave out mean for non-entrepreneurs or entrepreneurs on the scoring behavior of individual judges that are either entrepreneurs or non-entrepreneurs. Model 2 shows evidence of homophily: judges with investing background seem to pay more attention to the opinions of other judges with investing backgrounds and discount other judges. Model 4 still shows evidence of homophilous behavior by entrepreneurs in opinion formation. Model 6 reveals that lawyers pay less attention to themselves than they do to other committee members.

Tables and Figures

Table 1: Variable Definitions

Variable	Definition	Source
Startup Characteristics		
Female Founder	Dummy variable = 1 if at least one cofounder is female	MC, LI
Elite Edu	Dummy = 1 if at least one cofounder attended a	MC, LI
STEM	Dummy = 1 if at least one cofounder received a degree in engineering, science, or math	MC, LI
MBA	Dummy = 1 if at least one cofounder received an MBA	MC, LI
Ln Funding	Logged Thousands of Dollars of Outside investment received in first two years after potential graduation from MC program minus the measured treatment effect of the MC program for admitted firms	VX, CB
Judge Characteristics		
Investor	Dummy = 1 if judge is an investor	MC
Entrepreneur	Dummy = 1 if judge has started a company	MC, LI
Lawyer	Dummy = 1 if judge is a lawyer	MC, LI
Dependent Variables		
Judge Score	Score given by one judge	MC
Judge Score Difference	Difference of the	MC

MC – MC application; LI – LinkedIn; VX – VentureXpert; CB – Crunchbase

Table 2

Panel A: Summary Statistics

	Mean	Std. Dev	Min	Max
Startup Characteristics				
Female Founder	0.15	0.36	0	1
Elite Edu	0.21	0.41	0	1
STEM	0.26	0.44	0	1
MBA	0.15	0.36	0	1
Ln Funding	3.26	5.74	0	17.15
Judge Characteristics				
Female Judge	0.16	0.37	0	1
Investor	0.45	0.49	0	1
Entrepreneur	0.36	0.48	0	1
Lawyer	0.21	0.41	0	1
Round Characteristics				
# Judges, Paper Round	4.69	1.77	2	10
# Judges, Committee Round	5.07	0.87	2	9
Dependent Variables				
Judge Score	49.59	27.48	0	100
Avg. Judge Score, Paper Round	65.10	10.13	16	95
Avg. Judge Score, Committee Round	49.66	23.98	5	100

Panel B: Pairwise Correlations between Leave-out Means

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Leave out mean, investor	1.000									
(2) Leave out mean, not investor	0.307	1.000								
(3) Leave out mean, entrepreneur	0.671	0.502	1.000							
(4) Leave out mean, not entrepreneur	0.569	0.690	0.314	1.000						
(5) Leave out mean, lawyer	0.335	0.789	0.353	0.743	1.000					
(6) Leave out mean, not lawyer	0.840	0.496	0.796	0.603	0.395	1.000				
(7) Leave out mean, female judge	0.573	0.531	0.506	0.627	0.480	0.648	1.000			
(8) Leave out mean, not female judge	0.677	0.708	0.696	0.713	0.619	0.784	0.377	1.000		
(9) Leave out mean, high-quality judge	0.744	0.483	0.640	0.582	0.471	0.743	0.559	0.682	1.000	
(10) Leave out mean, not high-quality	0.555	0.750	0.586	0.682	0.658	0.670	0.581	0.718	0.365	1.000

Table 3: Relationship between Judge Value Add and Professional Background

	(1) Judge Value Add	(2) Judge Value Add	(3) Judge Value Add	(4) Judge Value Add
Entrepreneur	0.055			0.033
	(0.075)			(0.079)
Investor		-0.002		-0.043
		(0.074)		(0.080)
Lawyer			-0.104	-0.111
			(0.088)	(0.099)
Observations	371	371	371	371

Note: The regressions in this table are at the judge level. Each of the models is an OLS regression of each judge's computed value-added score on their professional background. Note that each judge can belong to multiple categories (e.g. one individual could be both an investor and entrepreneur). Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Social Influence in Committees by Judge Type

	(1) Judge Score	(2) Judge Score	(3) Judge Score	(4) Judge Score	(5) Judge Score	(6) Judge Score
Leave out mean	0.600***					
	(0.017)					
Leave out mean, investor		0.339***				
		(0.019)				
Leave out mean, not investor		0.251***				
		(0.016)				
Leave out mean, entrepreneur			0.310***			
			(0.019)			
Leave out mean, not entrepreneur			0.321***			
			(0.020)			
Leave out mean, lawyer				0.171***		
				(0.021)		
Leave out mean, not lawyer				0.427***		
				(0.025)		
Leave out mean, female judge					0.218***	
					(0.023)	
Leave out mean, not female					0.436***	
					(0.027)	
Leave out mean, high-quality judge						0.287***
						(0.018)
Leave out mean, not high-quality judge						0.280***
						(0.022)
Constant	25.362***	26.190***	23.310***	25.091***	21.812***	27.779***
	(1.054)	(1.169)	(1.278)	(1.405)	(1.498)	(1.324)
Observations	3615	3057	2590	2117	2713	2243
Judge Fixed Effects	Y	Y	Y	Y	Y	Y
F-Stat on Two Peer Effects		8.54	0.10	40.66	377.59	0.10

Note: These regressions are at the Judge X Startup level but now focus on a specific evaluation round. These regressions only include committees where there is at least one judge in leave-out mean with the professional background being studied in that model (e.g. investor or entrepreneur). Thus, the number of observations varies across the models. Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Consensus-Building in Committee Round by Judge Type with Focal Judge Interactions

	(2) Judge Score	(4) Judge Score	(6) Judge Score	(8) Judge Score
Focal Judge Type	Investor	Entrepreneur	Lawyer	High-Quality
Leave out mean, focal judge type	0.244***	0.200***	0.139***	0.175***
	(0.031)	(0.029)	(0.024)	(0.025)
Leave out mean, not focal judge type	0.408***	0.409***	0.477***	0.377***
	(0.032)	(0.033)	(0.029)	(0.035)
Judge is focal type X Leave out mean, focal judge type	0.104***	0.207***	0.100*	0.256***
	(0.040)	(0.040)	(0.056)	(0.038)
Judge is focal type X Leave out mean, not focal judge type	-0.215***	-0.116***	-0.174***	-0.116**
	(0.037)	(0.042)	(0.056)	(0.046)
Constant	25.365***	22.551***	24.662***	25.977***
	(1.174)	(1.280)	(1.422)	(1.351)
Observations	3057	2590	2117	2243
Judge Fixed Effects	Y	Y	Y	Y
Round	Committee	Committee	Committee	Committee

Note: These regressions are at the Judge X Startup level but now focus on a specific evaluation round. These regressions only include committees where there is at least one judge in leave-out mean with the professional background being studied in that model (e.g. investor or entrepreneur). Thus, the number of observations varies across the models. Standard errors in parentheses

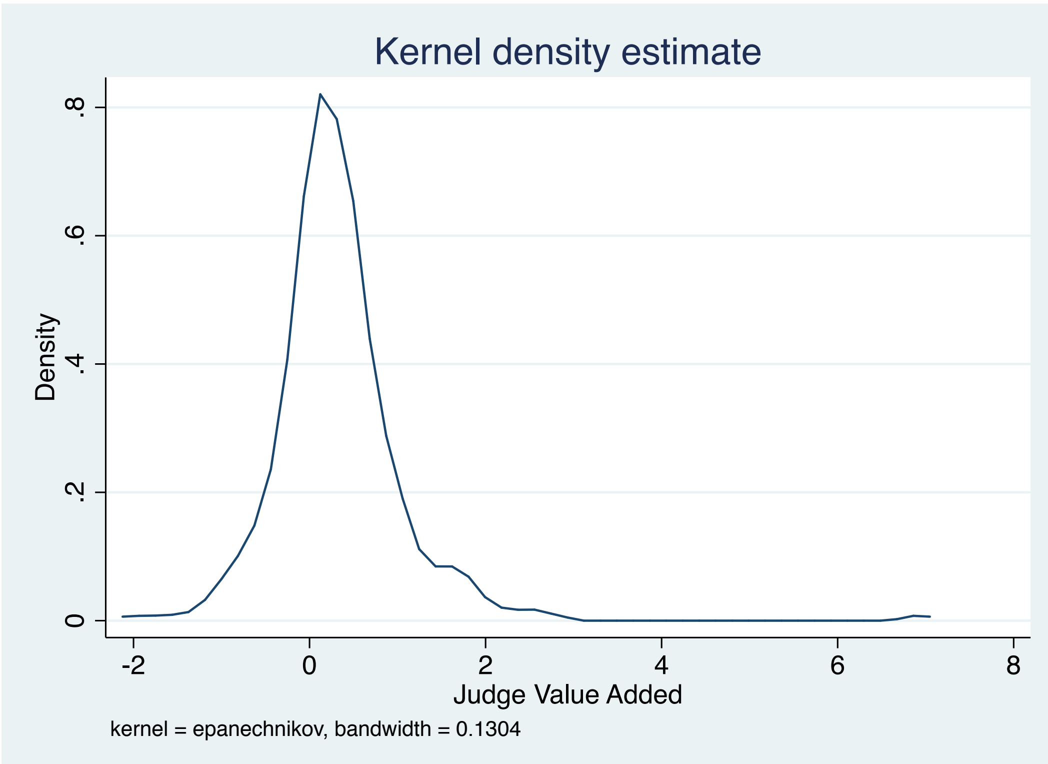
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Performance Impact of Consensus and Divergence of Opinion

	(1) Committee Average	(2) Committee Average	(3) Committee Average	(4) Committee Average	(5) Committee Average
Focal Judge Type	Investor	Entrepreneur	Lawyer	Female	High-Quality
Ln Funding, Below Mean Divergence	0.087	0.464^{***}	0.558^{***}	0.525^{***}	0.571^{***}
	(0.205)	(0.070)	(0.076)	(0.085)	(0.074)
Ln Funding, Above Mean Divergence	0.416^{***}	0.323^{***}	0.270^{***}	0.329^{***}	0.250^{***}
	(0.056)	(0.074)	(0.069)	(0.065)	(0.069)
Founders, STEM degree	0.915	0.867	1.089	0.853	0.794
	(0.826)	(0.826)	(0.824)	(0.826)	(0.822)
Founders, Grad Degree	2.631^{**}	2.649^{***}	2.620^{**}	2.631^{**}	2.576^{**}
	(1.022)	(1.022)	(1.018)	(1.021)	(1.017)
Founders, Elite Degree	2.693^{***}	2.741^{***}	2.579^{***}	2.698^{***}	2.724^{***}
	(0.882)	(0.882)	(0.879)	(0.881)	(0.877)
Founders, MBA Degree	0.683	0.615	0.665	0.658	0.629
	(1.053)	(1.053)	(1.048)	(1.052)	(1.047)
Founders, Prior Entrepreneurship	-1.123	-0.955	-1.156	-1.085	-1.122
	(0.871)	(0.872)	(0.867)	(0.869)	(0.865)
Founders, Female	-0.043	0.007	0.188	0.028	0.067
	(0.787)	(0.787)	(0.786)	(0.787)	(0.783)
Constant	58.155^{***}	58.269^{***}	58.269^{***}	58.214^{***}	58.266^{***}
	(0.564)	(0.562)	(0.562)	(0.564)	(0.561)
Observations	913	913	913	913	913
Committee FE	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y

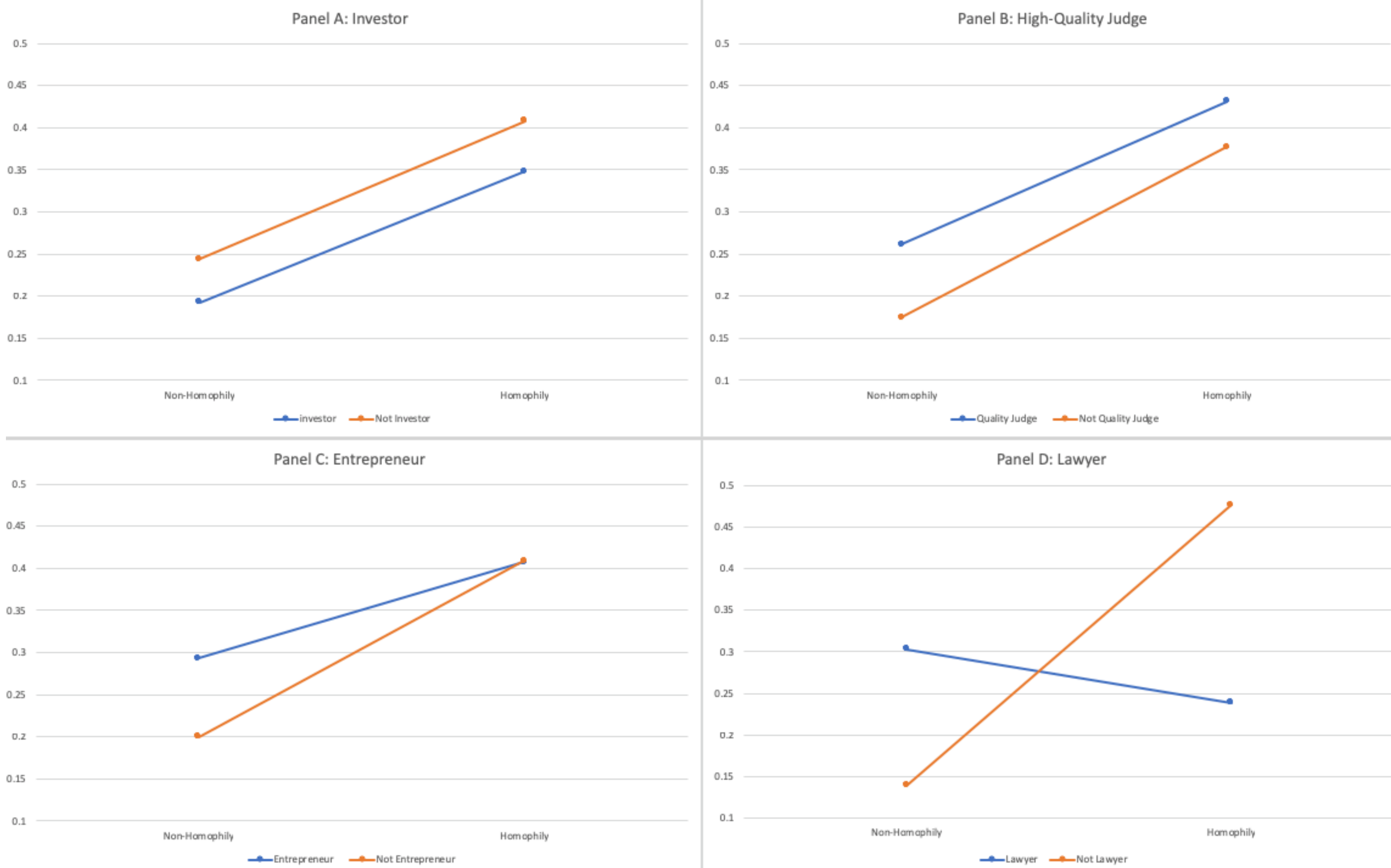
Note: These regressions are at the startup level and explore how above and below average levels of divergence between the scores of different groups at the committee level yield differences in the performance of the committees in evaluating startups. Each column represents a different focal group from the different professional backgrounds of the judges: investors, entrepreneurs, and lawyers as well as female judges and high-quality judges. We calculated the average score for the startup given by judges in and outside the focal group, then calculated the absolute value of the difference between the averages of the two groups. We then coded a dummy variable as either 1 or 0 depending on whether the committee had above or below average levels of divergence for that group category. We then estimated the model $Score_{ij} = \beta_j Q_{ij} + X_i + \epsilon_{ij}$ where Q_{ij} is the ex-post revealed quality of the startup (logged funding dollars two years after application minus the average treatment effect for admitted startups). The parameter β_j represents an estimate of the value add of committees with below and above average levels of opinion divergence.

Figure 1: Distribution of Judge Value Add



Note: This figure plots the distribution of Judge Value Added calculated using the method proposed in the working paper version of Li (2017). Using data at the Judge X Startup level, we estimated the model $Score_{ij} = \beta_j Q_{ij} + X_i + \epsilon_{ij}$ where Q_{ij} is the ex-post revealed quality of the startup (logged dollars two years after application minus the average treatment effect for admitted startups). The parameter β_j represents an estimate of the value add of each judge. The plot above reveals a right skewed distribution with mass centered just to the right of zero.

Figure 2: Plots of Combinations of Coefficient by Evaluator type from Table 6



Note: This figure attempts to make the results of Table 6 more easily interpreted. In particular, it plots the sums of the main effect and interaction effects for each of the models 6-2,6-4,6-6 and 6-8 depending on whether the individual judge in an observation is one with or without the professional experience that is being used for the interaction term in this model (e.g. Investor, Lawyer, etc) which we term the focal characteristic. For example, in model 6-2, an investor's individual score for any given startup is related to those of other investors (Homophily) with an additive parameter of 0.345 (0.244 +0.104) whereas the peer effect of non-investors on investors (Non-Homophily) is 0.193 (0.408-0.215).

References

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. "Some Simple Economics of Crowdfunding." *Innovation Policy and the Economy* 14 (January): 63–97. https://doi.org/10.1086/674021.

Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken. 2016. "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia." *American Economic Review* 106 (7): 1663–1704.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98–108.

Bayus, Barry L. 2013. "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community." *Management Science* 59 (1): 226–244.

Becker, Joshua, Devon Brackbill, and Damon Centola. 2017. "Network Dynamics of Social Influence in the Wisdom of Crowds." *Proceedings of the National Academy of Sciences* 114 (26): E5070–76. https://doi.org/10.1073/pnas.1615978114.

Blinder, Alan S., and John Morgan. 2000. "Are Two Heads Better Than One?: An Experimental Analysis of Group vs. Individual Decisionmaking." Working Paper 7909. National Bureau of Economic Research. http://www.nber.org/papers/w7909.

Botelho, Tristan L., and Mabel Abraham. 2017. "Pursuing Quality: How Search Costs and Uncertainty Magnify Gender-Based Double Standards in a Multistage Evaluation Process." *Administrative Science Quarterly* 62 (4): 698–730.

Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, and Christoph Riedl. 2016. "Looking across and Looking beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science." *Management Science* 62 (10): 2765–2783.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–882.

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76. https://doi.org/10.1257/jep.26.3.157.

Cialdini, Robert B. 1987. *Influence*. Vol. 3. A. Michel Port Harcourt.

Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annu. Rev. Psychol.* 55: 591–621.

Cohen, Susan, Daniel C. Fehder, Yael V. Hochberg, and Fiona Murray. 2019. "The Design of Startup Accelerators." *Research Policy*.

Criscuolo, Paola, Linus Dahlander, Thorsten Grohsjean, and Ammon Salter. 2017. "Evaluating Novelty: The Role of Panels in the Selection of R&D Projects." *Academy of Management Journal* 60 (2): 433–460.

Csaszar, Felipe A., and J. P. Eggers. 2013. "Organizational Decision Making: An Information Aggregation View." *Management Science* 59 (10): 2257–2277.

Cyert, Richard M., and James G. March. 1963. "A Behavioral Theory of the Firm." *Englewood Cliffs, NJ* 2 (4): 169–187.

Golub, Benjamin, and Matthew O. Jackson. 2010. "Naive Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics* 2 (1): 112–49.

———. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *The Quarterly Journal of Economics* 127 (3): 1287–1338.

Heath, Chip, and Rich Gonzalez. 1995. "Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making." *Organizational Behavior and Human Decision Processes* 61 (3): 305–326.

Hong, Lu, and Scott E. Page. 2004. “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers.” *Proceedings of the National Academy of Sciences* 101 (46): 16385–16389.

Huang, Laura, and Jone L. Pearce. 2015. “Managing the Unknowable The Effectiveness of Early-Stage Investor Gut Feel in Entrepreneurial Investment Decisions.” *Administrative Science Quarterly* 60 (4): 634–70. https://doi.org/10.1177/0001839215597270.

Kerr, William R., Josh Lerner, and Antoinette Schoar. 2011. “The Consequences of Entrepreneurial Finance: Evidence from Angel Financings.” *The Review of Financial Studies* 27 (1): 20–55.

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. 2014. “Entrepreneurship as Experimentation.” *The Journal of Economic Perspectives* 28 (3): 25–48.

Keuschnigg, Marc, and Christian Ganser. 2016. “Crowd Wisdom Relies on Agents’ Ability in Small Groups with a Voting Aggregation Rule.” *Management Science* 63 (3): 818–28. https://doi.org/10.1287/mnsc.2015.2364.

Kornish, Laura J., and Karl T. Ulrich. 2011. “Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas.” *Management Science* 57 (1): 107–128.

Li, Danielle. 2017. “Expertise versus Bias in Evaluation: Evidence from the NIH.” *American Economic Journal: Applied Economics* 9 (2): 60–92.

Li, Danielle, and Leila Agha. 2015. “Big Names or Big Ideas: Do Peer-Review Panels Select the Best Science Proposals?” *Science* 348 (6233): 434–438.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. “How Social Influence Can Undermine the Wisdom of Crowd Effect.” *Proceedings of the National Academy of Sciences* 108 (22): 9020–9025.

Mannes, Albert E., Richard P. Larrick, and Jack B. Soll. 2012. “The Social Psychology of the Wisdom of Crowds.” *Frontiers of Social Psychology. Social Judgment and Decision Making*.

Mollick, Ethan, and Ramana Nanda. 2015. “Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts.” *Management Science* 62 (6): 1533–1553.

Rao, Hayagreeva, Henrich R. Greve, and Gerald F. Davis. 2001. “Fool’s Gold: Social Proof in the Initiation and Abandonment of Coverage by Wall Street Analysts.” *Administrative Science Quarterly* 46 (3): 502–526.

Scott, Erin L., Pian Shu, and Roman Lubynsky. 2015. “Are’Better’Ideas More Likely to Succeed? An Empirical Analysis of Startup Evaluation.” *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, no. 16–013. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2638367.

Sethi, Rajiv, and Muhamet Yildiz. 2016. “Communication with Unknown Perspectives.” *Econometrica* 84 (6): 2029–2069.

Simon, Herbert A. 1957. “Models of Man; Social and Rational.”

Surowiecki, James. 2005. *The Wisdom of Crowds*. Anchor. https://books.google.com/books?hl=en&lr=&id=hHUsHOHqVzEC&oi=fnd&pg=PR11&ots=ZrcDXmOpfi&sig=QcvNr8x8Pb9FFgeXJwwr2sTov9k.

Teplitskiy, Misha, Hardeep Ranu, Gary Gray, Michael Menietti, Eva Guinan, and Karim Lakhani. 2019. “Do Experts Listen to Other Experts? Field Experimental Evidence from Scientific Peer Review.” HBS Working Paper 19–107.

Woolley, Anita Williams, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. “Evidence for a Collective Intelligence Factor in the Performance of Human Groups.” *Science* 330 (6004): 686–688.

Appendix A: Calculating Judge Value Add

Judge Value Added calculated using the method proposed in the working paper version of Li (2017). Using data at the Judge X Startup level, we estimated the model $Score_{ij} = \beta_j Q_{ij} + X_i + \epsilon_{ij}$ where Q_{ij} is the ex-post revealed quality of the startup (logged dollars two years after application minus the average treatment effect for admitted startups). The parameter β_j represents an estimate of the value add of each judge across the rounds. The plot above reveals a right skewed distribution with mass centered just to the right of zero.

Appendix B: Comparison of Social Influence across Different Judging Rounds

Table B1: Social Influence in Committees by Judge Type and Rounds

	(1) Judge Score	(2) Judge Score	(3) Judge Score	(4) Judge Score	(5) Judge Score	(6) Judge Score	(7) Judge Score	(8) Judge Score	(9) Judge Score	(10) Judge Score
Leave out mean	0.197***	0.600***								
	(0.021)	(0.017)								
Leave out mean, investor			0.105***	0.339***						
			(0.018)	(0.019)						
Leave out mean, not investor			0.082***	0.251***						
			(0.018)	(0.016)						
Leave out mean, entrepreneur					0.096***	0.310***				
					(0.019)	(0.019)				
Leave out mean, entrepreneur					0.097***	0.321***				
					(0.019)	(0.020)				
Leave out mean, lawyer							0.067***	0.171***		
							(0.020)	(0.021)		
Leave out mean, not lawyer							0.182***	0.427***		
							(0.027)	(0.025)		
Leave out mean, high-quality judge									0.093***	0.287***
									(0.021)	(0.018)
Leave out mean, not high-quality judge									0.045	0.280***
									(0.029)	(0.022)
Constant	52.971***	25.362***	53.829***	26.190***	53.021***	23.310***	48.834***	25.091***	56.764***	27.779***
	(1.391)	(1.054)	(1.548)	(1.169)	(1.687)	(1.278)	(1.968)	(1.405)	(2.151)	(1.324)
Observations	3681	3615	2965	3057	2666	2590	2102	2117	1624	2243
Judge Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Round	Paper	Committee	Paper	Committee	Paper	Committee	Paper	Committee	Paper	Committee

Note: These regressions are at the Judge X Startup level but now focus on a specific evaluation round. These regressions only include committees where there is at least one judge in leave-out mean with the professional background being studied in that model (e.g. investor or entrepreneur). Thus, the number of observations varies across the models. Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$