NBER WORKING PAPER SERIES

INFORMATION FRICTIONS AND EMPLOYEE SORTING BETWEEN STARTUPS

Kevin A. Bryan
Mitchell Hoffman
Amir Sariri

Working Paper 30449
http://www.nber.org/papers/w30449

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022, Revised July 2023

We thank David Autor, Shai Bernstein, Nick Bloom, Claudine Gartenberg, Matt Gentzkow, Jonah Rockoff, Kathryn Shaw, Chris Stanton, and especially Michèle Belot and Alan Benson for helpful comments. We are also grateful to numerous seminar and conference participants. We thank Daphné Baldassari for outstanding research assistance. Special thanks in memoriam to Dr. Peter Wittek for his kind assistance. We thank the anonymous science-based entrepreneurship program (SEP) for its collaboration and generous assistance in conducting the RCTs. The study and a pre-analysis plan were registered with the AEA RCT registry under ID AEARCTR-0004242. Our pre-analysis plan is viewable on the AEA RCT registry. We thank the Social Science Prediction Platform for their help in designing the economist expert survey. Two of the authors have performed paid work for SEP on topics unrelated to hiring and the workforce. Financial support from SSHRC and the Strategic Innovation Fund is gratefully acknowledged. We are deeply grateful to the Creative Destruction Lab for its support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information Frictions and Employee Sorting Between Startups
Kevin A. Bryan, Mitchell Hoffman, and Amir Sariri
NBER Working Paper No. 30449
September 2022, Revised July 2023
JEL No. M50,M51

ABSTRACT

Would workers apply to better firms if they were more informed about firm quality? Collaborating with 26 science-based startups, we create a custom job board and invite business school alumni to apply. The job board randomizes across applicants to show coarse expert ratings of all startups' science and/or business model quality. Making ratings visible strongly reallocates applications toward higher-rated firms. This reallocation holds restricting to high-quality workers. Treatments operate in part by shifting worker beliefs about firms' right-tail outcomes. Despite these benefits, workers make post-treatment bets indicating highly overoptimistic beliefs about startup success, suggesting a problem of broader informational deficits.

Kevin A. Bryan
University of Toronto
Canada
kevincure@gmail.com

Amir Sariri
Purdue University
403 W. State Street
West Lafayette, IN 47907
asaririk@gmail.com

Mitchell Hoffman
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON M5S 3E6
and NBER
mitchell.hoffman@rotman.utoronto.ca

A data appendix is available at http://www.nber.org/data-appendix/w30449

Hiring is a central issue in economics and especially in personnel economics (Ichniowski *et al.*, 1997; Oyer & Schaefer, 2011; Bloom & Van Reenen, 2011). As surveyed by Roberts & Shaw (2022), a growing body of work analyzes how firms select among workers, increasingly via randomized controlled trials (RCTs). Less is known about how workers choose among firms. Just as firms may have imperfect information about the skill of a potential employee, workers may have difficulty identifying "good jobs." Particularly for startups, it may be hard for workers to separate firms with promising futures from lemons.

Consider a worker choosing where to apply. With established firms, she might compare pay packages, consult employer review websites, and talk to contacts who have worked there. In contrast, small startups often pay similar low base salaries (Sorenson *et al.*, 2021), are unlikely to have online reviews, and have few past or present employees. This assessment process for applicants is even harder for *science-based* startups, such as those in machine learning or quantum computing. In these fields, the technology can be hard to evaluate, especially for non-experts, and market demand can be uncertain.

Workers with imperfect ability to evaluate firms may therefore apply to ones with limited potential. This harms workers and potentially economic efficiency more broadly if promising startups have trouble hiring good workers. While investors often conduct "deep diligence" to address information deficits before investing in startups, e.g., by paying outside medical school or computer science professors to evaluate the firms (Gompers *et al.*, 2020), potential employees presumably lack bargaining power to require such information.

How severe is this imperfect information to the functioning of labor markets, especially for startups? If credible information about firm quality were made available to workers, would they change who they apply to? Would they apply to *ex ante* better firms?

We address these questions using a pair of RCTs. In the *primary RCT*, we recruit 26 actively-hiring, early-stage startups from a world-leading science-based entrepreneurship program (SEP) with over $20 billion in equity from the first ten cohorts of its participating firms (described more in Section 1). After these startups provide us with job advertisements, we build a custom job board accessible to nearly 20,000 business school alumni. In addition to the startups' own ads, subsets of the applicants are randomly provided coarse ratings of each firm's science quality from leading scientists and/or ratings about the firms' business model quality from experienced incubator staff. As in other job application settings, applicants are free to use firm websites, press coverage, and so on to investigate firms they consider applying to. In total, 1,877 applications are made by 250 job-seekers.

None of the applicants are aware of the information randomization. From their perspective, the job board looks similar to other recruitment websites. Treated workers are exposed to expert ratings for all firms, allowing us to examine the impact on workers of

market-level shifts in the precision of information (e.g., how would workers respond if the government or a jobs platform provided expert ratings on all startups in addition to startups' own job advertisements). The *secondary RCT* examines similar choices to the primary RCT, but in a highly controlled environment. In it, 191 MBA students examine several real startups and answer incentive-compatible belief questions (explained further below), just as in the primary RCT, but they state their hypothetical interest in working at the startups after graduation instead of making actual job applications.

It is not *ex ante* theoretically obvious that applicants would react to expert ratings of science or business quality. They will not react if they do not care about these features—perhaps they apply to all firms in a given industry in a given city paying a given salary. They will also not react if they already have precise information about these dimensions of firm quality.[1] Expert ratings are predictive of firm success, both in past data and for the firms in the RCT.

Our paper's main finding is that expert ratings substantially affect what firms applicants apply to. Both science and business quality ratings matter and both matter to a broadly similar degree. Relative to not providing information, providing positive information on science quality increases the probability a worker applies to a given firm by 12%, while negative information decreases application probability by 24%. Likewise, a positive signal about business model quality raises application probability by 29%, and negative information decreases applications by 12%. Put another way, startups with above average business and science quality receive 11% more applications than those below average on each dimension when workers receive no additional signals. However, in the treatment where workers receive both quality signals, that gap increases to 80%.

To better understand these effects on worker preferences, we examine how the treatments change worker beliefs. Expert ratings substantially affect workers' perceptions of the science and business quality of firms. They also affect workers' beliefs about whether firms will succeed. To ensure that workers provide thoughtful answers regarding firm success, we incentivized worker beliefs using a betting game, developed in experimental economics, where participants could win up to $250 CAD (≈ $200 USD). The betting game incentivizes participants to provide honest beliefs about the probability observable events will occur. Even under significant incentives, workers are overoptimistic about firm success, and they dramatically overestimate the chance that firms will achieve a successful exit or IPO within a year. Despite this, workers update their beliefs in response to expert ratings, particularly

[1]Business and science quality are not the same. Firms can be good at one but not the other. Consider Theranos, a health startup once valued at over $8 billion. There was huge demand for Theranos' product as marketed by its CEO, but the underlying science was poor (Carreyrou, 2018). In our data, startups' science and business quality scores are uncorrelated, as detailed below in Section 1.1.

beliefs about whether firms will raise venture capital. These results suggest that beliefs about firm quality and success are a mechanism for our paper's main finding.

Turning to treatment effect heterogeneity, we show there is limited heterogeneity for most dimensions of worker and firm quality. There is no significant heterogeneity for worker startup experience or STEM background, though we find that men respond more than women to science quality ratings. A critical dimension of heterogeneity in theory is worker quality, which we measure by having a startup-focused HR expert rate resumes. If anything, higher-quality workers respond more strongly to our intervention than lower-quality workers. Importantly, there is no evidence that our overall treatment effects are driven by low-quality workers, suggesting that interventions like ours would not simply drive low-quality workers to high-quality firms.

Returning to whether the results were obvious, following DellaVigna & Pope (2018), we asked economist experts in related fields to predict the main results of the primary RCT in terms of job applications. Economists correctly predict that providing expert ratings would affect job applications, but most economists fail to predict the qualitative features of the effect. Economists predict that negative information will have a larger effect on job applications, though we find positive and negative information have effects of similar magnitudes. Most economists incorrectly believe that the treatment effect will vary by STEM degree and worker quality, but fail to predict heterogeneity by gender. Quantitatively, on average, economists underpredict the impact of expert ratings in driving applications to highly rated firms, though there is wide heterogeneity in predictions.

Our paper contributes to four literatures. First, it contributes to work in personnel economics, organizational economics, and labor economics on worker/firm matching (Bandiera *et al.*, 2015), where a growing body of research uses natural experiments or RCTs to understand how workers choose among jobs or firms. To our knowledge, our paper is the first RCT in this literature to study how workers choose between startups, as well as the first about the role of imperfect information about firm quality in affecting how workers choose among firms.[2] We provide the first evidence that even highly-educated workers have limited ability to identify high-quality startups. Broadly speaking, our results also speak to understanding how persistent differences across firms in performance (Ichniowski *et al.*, 1997; Bloom & Van Reenen, 2007; Syverson, 2011; Bloom *et al.*, 2013, 2019) are appreciated by workers.

[2]Other recent studies using natural experiments or RCTs to understand choice among jobs or firms include Ashraf *et al.* (2020); Flory *et al.* (2015); Hedegaard & Tyran (2018); Stern (2004); Wiswall & Zafar (2015). Unlike our paper, other papers in this literature generally focus on established firms or all firms, and analyze other characteristics like whether a firm is family-friendly or lets scientists publish. Benson *et al.* (2020) create firms on mTurk and show that randomly endowed better reputations increase job fill rates. In finance, Bernstein *et al.* (2022) show that workers become more interested in startups where a prominent venture capitalist has invested. Appendix B.1 discusses more on related work outside of economics.

Our results provide the first evidence that workers would apply to substantially different firms if they had more precise knowledge about the science or business model quality of firms. Policies that provide credible information about firms to workers, whether by governments, business councils, entrepreneurship programs, job boards, or firms themselves, can improve worker welfare and reduce potential misallocation of workers to startups. Just as *physical capital* may be misallocated due to frictions between firms, so too may *human capital* (Hsieh *et al.*, 2019). Of particular relation to our paper is the work of Belot *et al.* (2018, 2022a), who also conduct RCTs where information about aspects of jobs is randomly provided to jobseekers. Belot *et al.* (2018, 2022a) show that providing generally available labor market data to unemployed jobseekers, such as skill requirements for different jobs, has sizable effects on their job applications. Our results indicate that informational deficits exist in labor markets even for highly skilled and advantaged workers.[3]

Second, our results relate to work in personnel economics on startups. A central question is why startups and other firms often pay workers with equity instead of salary, common answers being taxes, credit constraints, or aligning worker and firm beliefs (Oyer, 2004; Oyer & Schaefer, 2005). Bergman & Jenter (2007) propose an alternative: workers overestimate the probability of positive events occurring to workers. To our knowledge, our paper presents the first direct evidence, obtained using incentivized experimental methods, that workers overestimate the probability of a successful startup exit.[4] Thus, our work suggests the possibility that firms may find it cheaper in expected value to pay workers in equity instead of salary. Moreover, since workers respond strongly to expert ratings, workers may face challenges in accurately evaluating expected returns from equity-based compensation in the absence of expert ratings. Our results bear on recent discussion by business and legal scholars about whether additional regulation should be considered regarding startup compensation toward employees (Aran & Murciano-Goroff, 2021).

Third, our paper contributes to work in behavioral labor economics. It has been shown that workers exhibit behavioral tendencies in the context of job search, including present bias (DellaVigna & Paserman, 2005; Belot *et al.*, 2021), reference dependence (DellaVigna *et al.*, 2017), and overconfidence (Spinnewijn, 2015). Our paper shows that workers also exhibit overoptimism, believing that firms will be more successful than they actually are.

Fourth, our paper contributes to work in entrepreneurship regarding resource acquisi-

[3]Belot *et al.* (2018, 2022a) study information frictions about what types of jobs are available or would be a good fit, whereas we study information frictions in regard to firm quality. Dustmann *et al.* (2022) show that a non-information policy, higher minimum wages, reallocates workers to better firms. Jäger *et al.* (2022) show that workers have biased beliefs about their outside option in the broader German labor market. There are non-job parallels of RCTs addressing information frictions in choices over schools (Hastings & Weinstein, 2008), drug plans (Kling *et al.*, 2012), and neighborhoods (Bergman *et al.*, 2019).

[4]There is substantial work on CEO overconfidence, but our focus is overconfidence of lower-level workers.

tion by startups. Past work examines why good startups have trouble acquiring resources such as capital and partnerships (Lerner, 1995; Hsu & Ziedonis, 2013). Our work helps address why good startups have difficulty acquiring workers, and suggests that good startups may have trouble hiring because workers do not know which startups are most promising.

1 Context, RCT Design, and Theoretical Framework

Context. SEP is a non-profit, selective, nine-month program operating across several high-profile business schools in six countries. The core mission is to provide mentorship to early-stage, science-based startups. This program is akin to a forum that convenes angel and institutional investors and technical experts five times a year to offer structured mentorship to startups.[5] At the time of our RCTs, about 900 ventures had gone through the program, including 389 in the 2018-2019 cohort, indicating the rapid growth of the program.

SEP describes its program as suitable for *seed-stage* startups, meaning ventures expecting to raise capital during or after the conclusion of the program. Ventures in other stages of their financing life-cycle may still be admitted if they are expected to benefit from the program. However, ventures that are believed not to be fundable or scalable are not usually admitted to the program. As the program progresses through each of its five meetings, a subset of the startups gets cut from the program. The average graduation rate over the first seven cohorts of the program is approximately 40%.

To support startups with distinct technological and growth trajectories, SEP is offered through specialized streams such as machine learning, quantum machine learning, space, health, and energy. Streams involve staff and mentors with experience or expertise in the corresponding field or market. For example, the space stream's mentors include three astronauts, leaders of public and private space exploration entities, and investors in the launch and propulsion sectors.

There are three features of SEP that make it an ideal setting for our research questions. First, SEP is one of the largest and most highly-esteemed programs of its kind in the world. SEP's stature means it gets access to world-class science and business expert raters, and also that we have a sizable number of startups who are interested in participating in the RCT.

Second, SEP startups are the type of startups who frequently engage in hiring. SEP is not a program for student companies or projects. Rather, all firms in SEP are existing startups, many of which include world-leading scientists on their founding team, who are

[5]SEP's mentorship structure is aimed at helping startups design and prioritize short-term measurable objectives. For brevity, we avoid describing aspects of SEP that are not relevant to our study.

thought to possess a chance of becoming a large, venture-backed business.[6]

Third, ours is a highly natural setting for thinking about uncertainty over business and science quality since the firms we study deal in cutting-edge science and business problems. This is an important feature of our setting given our research question, though we also fully acknowledge that our paper's findings do not necessarily apply to all young firms (e.g., to firms dealing with simple consumer problems).

1.1 Primary RCT: Job Board

Origins of the RCT and firm recruitment. Hiring is a natural part of growth and development for SEP firms, and one that firms often struggle with anecdotally, even when they are high-quality. In 2019, SEP decided to launch a pilot job board, which we helped design and implement. A key factor for SEP to engage in the RCT was a belief that its strongest firms were having a hard time hiring, perhaps because it was hard for workers to identify their quality. Past research has argued that startups may face challenges competing with established firms for talent (Manchester *et al.*, 2023), and this is also true in our setting. However, SEP was especially concerned that job applicants face particular challenges in distinguishing between startups and this helped motivate SEP's desire to conduct the RCT.

SEP emailed the 183 firms who were part of the 2018-2019 cohort of firms that participated in the program at SEP's headquarter location, and asked if they wished to participate. Firms were told truthfully that experimentation may occur, but were not informed specifically about the nature of the RCT. Of the 183 firms, 26 firms (or 14%) agreed to participate. As analyzed later in Section 2, participating firms seem broadly representative of SEP firms. Each startup wrote short advertisements about their firm. Like most startups, SEP firms tend to pay relatively low wages, but offer equity to new hires.

In creating the job board, a key goal was to make it similar to existing job boards for startup firms (e.g., AngelList careers, or Y Combinator's `workatastartup.com`). Specifically, SEP wanted its job board to be easy to use, visually appealing, and feature the type of information that would appear on AngelList. Firms were assigned positions on the board alphabetically from left-to-right and top-to-bottom based on the founders' first name.[7] Figure 1 shows the type of information workers saw when they clicked on a firm logo.

[6] To fix ideas, a reader might think of a typical SEP startup as consisting of two computer science professors with an advancement in selecting images for machine learning by autonomous vehicles, or of a doctor and scientist with a new application of machine learning to healthcare problem. This is distinct from student startups, where time commitments are limited and who are less likely to grow and hire people in the future.

[7] It was not possible logistically to randomize the visual position of firms across workers. This is not an important concern for us because our results rely on within-firm differences across users in whether ratings were provided about a firm.

Worker recruitment. We recruited applicants for the RCT by partnering with two prominent North American business schools. Each agreed to email their alumni list customized and trackable links. The alumni emailed were graduates from MBA, specialized masters, and undergraduate business programs.[8]

Experimental procedure. Each potential applicant arrived at the job board via a code hidden in the website URL which triggered a randomized change in the visibility of two aspects of information about each firm, for four total arms. When a potential applicant viewed details of a particular firm, they would see, in addition to a link to the firm's website and a self-written description of the firm, one of four information treatments: a control, information about business model quality, information about science quality, or both. The treatment was assigned at the applicant level, so a given applicant would either see this information for all firms, or for none. Applicants were unaware that other applicants may have seen different information.

Since job seekers accessed the job board via a custom link, we can match the randomization each received to their name and ensure that the randomization treatment they receive stays constant in case they visit the board multiple times. Randomization in the emailed links was stratified by gender, graduation year, and current city, where known. The job board is a standard website viewable in any browser, so there was no restriction on their ability to search for further information about any firm, or even attempt to contact firms before applying.

At any point after investigating these firms, over a roughly four-week period, workers could upload their resume to the centralized job board application system. Applicants were told that a small number of resumes would be highlighted in emails to each participating firm. The stated motivation was that, since these firms are small, the SEP did not want to overburden the founders with hundreds of resumes. In addition, and for the same stated reason, applicants were only permitted to apply to up to ten startups each, and were asked to rank their relative interest in each firm.

Applicants were told that the likelihood of having their application sent to a firm was higher for firms they ranked higher, and further, that the names would be chosen using an incentive-compatible mechanism: random serial dictatorship (RSD) (Abdulkadiroglu & Sonmez, 1998). Under RSD, workers are first placed in a random order, and firms are given a fixed number of "slots". Since workers can rank up to ten startups, there are up to ten

[8]From the first business school, we contacted 5,681 undergraduate alum (graduated from 2008 to 2019) and 7,894 graduate alumni (graduated from 1946 to 2018). From the second business school, we contacted 3,701 undergrad alumni (graduated from 2009 to 2015) and 2,083 grad alumni (from 2009 to 2019). PhD alum were not emailed. Further details are in Appendix D.1.

rounds. In each round, the algorithm then goes through workers in order, matching workers with their highest-ranked firm which still has remaining slots. This is analogous to the draft in the National Basketball Association where draft order is random and the number of slots (i.e., teams per player) is one. Although we left details of RSD in the experiment to a linked text, we explained that an award-winning algorithm ensured that it was in the best interest of candidates to truthfully rank firms in order of their interest:[9]

> "To avoid inundating these start-ups with an excessive number of resumes, we have agreed to forward a limited number of resumes to each startup. **It is in your interest to state your true preference ranking!** Specifically, the probability your information is sent to a given venture is strictly higher the higher you rank a venture. An algorithm by leading economists ensures that there is no benefit to manipulating your true preference about which ventures you would like to meet." [see Appendix E for screenshots of how this appeared to subjects]

The limit of ten applications imposes a "cost" on applicants via the opportunity cost of not signaling interest in other firms. This cost is meant to mimic the actual cost of applying for jobs in a less centralized job search environment.[10] As described in Section 2 below, we investigate four pre-registered outcomes using these rankings: the probability of applying at all, the probability of listing a firm in one's top N ranks (we use top rank and top three rank), and the absolute rank of the firm (with unranked firms treated as having been ranked $N+1$). As seen below, results are similar across all four outcomes, including top rank and top three rank, which are not affected by the ten application constraint, suggesting that the constraint does not drive our findings.

Worker beliefs. After submitting their ranked list of firms they wished to apply to, job seekers were asked to voluntarily evaluate three randomly selected ventures from the set of firms they had just considered applying to. This request was made only after the ranked true applications were sent, to limit Hawthorne effects. For each job seeker, the three firms

[9] Past work indicates that RSD often yields incentive-compatible choices (Chen & Sönmez, 2002; Artemov *et al.*, 2017), and we expect that our high-skill subjects would have an easier time understanding it than most. Note that RSD is incentive-compatible for interest in *getting an interview*, not in having a job, since workers may worry that ex-ante better firms will get better applicants, and hence the probability of getting a job conditional on getting an interview is lower. We view this fact as a feature and not a bug of our setup, as the same issue would be present in a full-scale policy rollout where expert ratings or other signals were provided to all workers. RSD is used around the world in many labor markets (Fadlon *et al.*, 2022).

[10] On many job boards, applicants are unrestricted in the number of firms they can apply to, but each application costs time due to different application formats. However, some job boards and job matching processes restrict the number of applications, e.g., the National Resident Matching Program for US physician residents imposes extra fees for applying to more than 20 programs and forbids applying to more than 300. Our anecdotal sense is that firms and workers in the RCT saw the 10-firm constraint as very natural.

in question remained the same for all belief questions. We asked candidates to estimate each firm's science quality, business quality, probability of raising capital at a valuation of $1 million or more within a year, probability of having an IPO or being acquired for at least $50 million within one year, and their interest in working for the company. Before answering the two questions on probabilities (i.e., the IPO and capital raise questions), subjects were given a standard "explanation of probabilities" as developed in the work of Charles Manski, and as used in many subsequent papers and large-scale surveys (see the review by Bruine de Bruin *et al.* (2023)). The explanation gives examples of probabilities, and is designed to be simple and avoid leading subjects in any way (see Appendix E for the exact wording).

The probability questions were incentivized with a risk-invariant quadratic scoring rule (McKelvey & Page, 1990), under which applicants can win up to $250 CAD (≈ $200 USD) on the basis of the accuracy of their predictions.[11] Workers' perceptions of startups' science and business quality cannot be incentivized because they are subjective.

Science score. To grade the quality of each startup's underlying science, we use detailed scientific evaluations that SEP conducts to assist intake into each cohort. These evaluations are otherwise not made public. The assessment is performed by a group of distinguished university scientists and research scientists from Canada's National Research Council. The assessments are based on a 30-minute interview with the venture by a scientist with expertise in their core technology (e.g., a machine learning startup will be evaluated by a scientist with expertise in machine learning), as well as detailed written materials the venture provided in advance of the meeting. The scientists are paid to do the assessments as part of their regular salary from the National Research Council, and thus, take the assessments quite seriously.[12]

When onboarding scientists to conduct the assessments, SEP tells scientists to explicitly focus on the viability of the core technology and the technical ability of the founders to execute, regardless of what they think about the market potential or business viability of the technology. In the RCT, ventures on the job board are divided into above- and below-median for their science quality. A binary version of the rating was used to make the ratings easy

[11] In greater detail, if a subject guesses correctly and states confidence level c, they get a lottery with a $2c - c^2$ probability of winning $250 CAD and a $(1-c)^2$ probability of receiving zero. If they guess incorrectly, they get a $1 - c^2$ probability of winning $250 CAD and a c^2 probability of receiving zero. Under these incentives, it is optimal for subjects to accurately report their true confidence. Following past applied work (Hoffman, 2016), we explained that this system made it optimal for people to state their true beliefs, and then we provided the mathematical formulas separately for people to look at if interested. Appendix B.2 provides further discussion of past evidence supporting the reliability of eliciting beliefs in this manner.

[12] The National Research Council is the main scientific and technology research institute in the Canadian federal government. The SEP and National Research Council have an agreement for the scientists to have their work doing assessments count toward their salaries.

to interpret for jobseekers, and coarse ratings are common for many ratings (e.g., pass/fail for health inspections).[13]

Business model score. The business quality score is performed by full-time staff from the SEP who specialize in evaluating startups.[14] Like the science score, the business model score is based on an in-person interview and extensive supporting documents. The score is normally assigned at the time the firms apply to SEP. However, since the job board occurred a year after the business model was first evaluated, and firms often update their business model, SEP staff re-evaluated the business model score. They did so based on their interactions and conversations with the firms about the business model. SEP staff evaluate business quality along three margins: size of the market being targeted, quality of the business model, and ability of the team to execute on this opportunity. These three margins are considered critical aspects of a firm by venture capitalists and are a standard way of measuring the potential of a startup (Gompers *et al.*, 2020). We average these three scores to create the overall business model score.

The evaluators were unaware which aspect of their evaluation was being used, or the purpose of the re-evaluation. However, like the scientists, we believe that the business experts took the scores very seriously, as evaluating startup business models is a central part of their job at SEP, and they have career incentives to produce thorough and accurate ratings. Once scores were collected, ventures on the job board were divided into above- and below-median business model scores, dichotomized to be easy to interpret for jobseekers.

Critically, the business model score does not include any evaluation of the firm's underlying science. The evaluators are not PhD scientists and are told to focus solely on evaluating firm business models. Business model scores are uncorrelated with science scores for our 26 firms, both in continuous form (correlation coefficient of $\rho = 0.06$, which is indistinguishable from zero with $p = 0.75$) and when they are dichotomized ($\rho = -0.23$, $p = 0.27$).

Expert score predictiveness. To what extent are expert ratings predictive of actual firm success? We consider this question both in terms of historical data from SEP and for the 26 startups in our primary RCT, and find evidence of predictive power in both.

While the firms in the RCT are both relatively small in number and recently founded,

[13] The number of scores above/below median is not perfectly even because the underlying score is discrete (1-5 for science, 1-10 for business). For ethical reasons of protecting the ventures, SEP required that the information be presented to subjects as "Science Quality was Rated as Above Average: Yes" or "Science Quality was Rated as Above Average: No." This phrasing leaves some ambiguity about the score and is likely to be a more policy-relevant way of presenting scores in other contexts.

[14] These staff members are analogous to the portfolio managers at venture capital firms who specialize in narrowing the list of potential investments for senior partners. In fact, it is common for SEP staff to go work for venture capital firms as portfolio managers.

we exploit the fact that science ratings have been conducted on SEP firms for several years. Table 2 examines the correlation between expert science ratings and two outcomes, namely, (1) whether the firm graduated from the SEP program and (2) whether the firm raised money after the SEP program. As seen in Panel B, a 1σ increase in science rating predicts that a firm will have a 9.4 percentage point higher chance of raising money after the SEP, off an overall mean of 25%. Likewise, a 1σ increase in science rating predicts a 9.2pp higher likelihood of graduating from the SEP program, off a mean of 41%.

Turning to the firms in our RCT, of our 26 ventures, by August 2022 (i.e., three years after the start of the RCT), 17 were still in business ("survival"), 8 had publicly raised an additional $4 million USD or more, and a further 3 had hired at least 10 employees ("raised or hired"). Both business and science ratings are predictive of these outcomes: an above-average business model rating increases the probability of survival from 50% to 79%, while an above-average science rating increases it from 64% to 67%. More starkly, an above-average business model rating increases the probability of "raised or hired" from 42% to 50%, and an above-average science rating increases "raised or hired" from 29% to 58%.

Remarks on the design. Note the most important aspects of this information treatment. First, the business model and science evaluations were performed by experts in the respective domains, using information that goes well beyond what would generally be found on a company website. Second, above-average and below-average in the information treatments were relative to the selection of companies on the job board. These ventures, just by virtue of having taken part in the SEP, are already well in the upper tail of quality of all tech-based startups. Third, workers are shown truthful information at all times, where we vary the amount of information shown to the worker. That is, the information treatment is a coarse signal (a binary above/below average rating) where we either show the true binary score or not. This is a contrasting approach to an audit study where participants receive the same amount of information but certain information elements are randomly varied.

Finally, the information treatment here is analogous to treatments that policymakers, incubators, or job search websites could pursue. For instance, a logo denoting firms in an incubator that were thought to have the best underlying science, or a particularly promising business model, could be added to the incubator's employment website. Job search websites, or startups themselves, could explicitly highlight competitive markers of quality such as participation in a top incubator, investments from prominent venture capitalists, or the scientific renown of the founding team. In the discussion of our results, we give evidence about the extent to which this currently happens.

Timing and implementation. The job board started in May 2019. Emails were sent

out in 3 batches (batch 1 = MBA alumni of the 1st business school, batch 2 = undergrad alumni of the 1st business school, batch 3 = MBA and undergrad alumni of the 2nd business school), as detailed in Appendix D. For each batch, the job board was active for about one month.

After the application period ended, startups were emailed a link with secure access to the resumes of applicants who used the job board. They also received the name of ten candidates who showed particular interest in the venture, as measured by the RSD mechanism (see Appendix B.3 for further details on RSD implementation). Four months after the job board closed, we followed up with ventures about interviews or hires made on the basis of these applications. We then followed the startups through August 2022 to track their outcomes.

In total, 250 workers applied to at least one firm, and 1,877 total applications were submitted (i.e., the firms were ranked by an applicant). Most workers applied to between 5 and 10 firms. Time stamp data suggest that candidates took the process seriously. As seen in panel (a) of Figure 1, jobseekers land on a webpage where they can browse the different firms, as well as enter their contact information and job application rankings. After jobseekers first click on the data entry part of the webpage (which is presumably after they have started browsing the firms on the job board), the median time spent on the job board and answering the beliefs questions was 22 minutes (25th percentile was 12 minutes and 75th percentile was 54 minutes).

1.2 Secondary RCT: MBA Student Experiment

In March 2018, we conducted a secondary RCT with MBA students applying for competitive entry to a course associated with the SEP. This allows us to perform similar analyses to the primary job board RCT, but under highly controlled conditions that minimize inattention. As part of admissions to the SEP MBA course, candidates were asked to evaluate 3 randomly chosen firms among firms that had recently participated in the SEP. To do so, they received corporate information about 3 firms from a set of 20, including descriptions of the firm's product, founding team, and business strategy, plus technical briefing documents. Figure 2, Panel A, shows an example. For each firm, MBA students fill out a quantitative evaluation and answer some qualitative questions on suitability for SEP.

This evaluation was performed in a controlled classroom environment. Students had 40 minutes to evaluate each firm, and most students took most of the full 2 hours. Students were also told that their responses to this evaluation, particularly the qualitative questions not part of our study, would determine whether they were accepted into the SEP MBA

course. Thus, students took it seriously. As in the primary RCT, students were given firm documents that were randomized to either include no additional information, a binary expert evaluation of the firm's science, a binary expert evaluation of their business model, or both, and within student, all firms are treated the same (e.g., one sees a business expert rating for all firms). The source of these evaluations was identical to the primary RCT, although the firms in question were not identical. Figure 2, panel B shows how expert ratings were presented.

Instead of providing a ranked ordering of firms using incentive-compatible RSD, students were asked how interested they would be working in the firm after graduation on a 1-5 scale.[15] We also elicited the same beliefs about firm outcomes and the perceived quality of science and business as in the primary RCT. This allows us to analyze worker beliefs using pooled data from both RCTs.

1.3 Theoretical Framework

In Appendix C, we provide a simple equilibrium model of costly job application where workers observe the quality of firms imperfectly when deciding whether to apply. For some fraction of firms, workers are unable to separate productive from less productive ones. Reducing this imperfect information via expert ratings increases applications to above-average firms and decreases applications to below-average firms. This prediction holds even when jobseekers are forward-looking about competing with one another for jobs at better firms. Although our actual RCT was designed and powered to study applications instead of hiring outcomes, we show in the model that expert ratings lead better firms to attract higher-quality hires.

Ultimately, whether expert ratings affect job applications and beliefs about firm success and quality is an empirical question. Also, it is not clear which dimensions of startups (science or business quality) matter. We turn to these next.

2 Outcomes, Empirical Strategy, and Randomization

Outcomes. Via the RSD mechanism, our RCT uses applicant rankings to determine which job applications are highlighted to firms. Thus, we measure worker interest across firms using these rankings. In our pre-analysis plan, we specified that we would consistently analyze four different functional forms:

[15] Job applications are the key object of our study. Because expressed interest in the secondary RCT cannot be incentivized, we restrict our main results on application behavior to the primary RCT. Results using the non-incentivized job applications from the secondary RCT yield similar conclusions.

1. Whether a job candidate ranked a firm at all. We refer to this generically as whether someone applied to a firm.

2. Whether a firm was a candidate's top choice.

3. Whether a firm was in a candidate's top 3 choices.

4. The normalized rank of a firm. Specifically, a top ranked firm received a score of 10, the second rank firm a score of 9, ..., the 10th rank firm a score of 1, and unranked firms a score of 0. We then normalized this score.

Job applications are a central object of interest in personnel economics and provide the cleanest expression of jobseeker preferences. We discuss the subsequent outcomes of interviews and hiring in Section 3.2.

Empirical strategy. As pre-registered, we run regressions of the following form:

$$y_{nf} = \alpha_0 + \alpha_1 \text{GotBizInfo}_n + \alpha_2 \text{GotBizInfo}_n \times \text{GoodBizFirm}_f + \boldsymbol{X}_{nf} + \varepsilon_{nf}$$
$$y_{nf} = b_0 + b_1 \text{GotScienceInfo}_n + b_2 \text{GotScienceInfo}_n \times \text{GoodScienceFirm}_f + \boldsymbol{X}_{nf} + \varepsilon_{nf}$$

Here, n denotes workers and f denotes a firm that a worker is evaluating. Thus, an observation is a worker-firm. The outcome, y_{nf}, will be one of the above outcomes. The regressor GotBizInfo_n measures whether subject n is randomly assigned to receive information about the business quality of the firm. Likewise, GotScienceInfo_n measures whether a subject randomly receives information about the science quality of the firm. The variable GoodBizFirm_f indicates whether firm f is rated positively or not in terms of business quality, whereas GoodScienceFirm_f indicates whether a firm is rated positively or not in terms of science quality.

The controls $\boldsymbol{X}_{nf}$ will include firm fixed effects to control for the underlying quality of the firm, as well as any strata dummies for RCTs when we do a stratified randomization. We cluster the standard errors by worker to account for the fact that the treatments are at the worker level.

Our pre-analysis plan (PAP) also indicated that worker beliefs and perceptions of firm quality would be secondary outcomes of interest. These are analyzed in the same way as our results on worker application rankings. Finally, the PAP specifies that we will run heterogeneity analyses based on whether workers have a STEM degree or not.[16]

[16] In addition, the PAP states that we will try to perform heterogeneity analysis based on characteristics of the firms related to the degree of technological sophistication. In practice, the vast majority of the firms are highly sophisticated technologically, and it is not obvious how to compare the sophistication of, e.g., an AI fintech company to a quantum drug discovery platform.

Randomization check. Table 1 shows that the four treatment groups of the Primary RCT are balanced on observable characteristics. The Secondary RCT's randomization was not stratified, but we found the participants to be balanced on race. Note that the characteristics in Panel A refer only to workers who applied to at least one firm, for which we can observe their resume. A strong majority of applicants are currently employed, reflecting that this is a high-skill group. Workers have an average of about 10 years of experience and nearly half have an undergraduate STEM degree.

Panel A shows that the number of workers who apply differs slightly by treatment group. This reflects that we sent emails to equal numbers of business school alumni across the treatment groups, but the actual number who apply can still vary. There are no statistically significant differences across treatment arms in terms of applicant location, gender, graduation year, startup work history, years of work experience, or STEM background.

Worker selection into the RCT. Of people emailed from our partner business school alumni lists, 1.33% participate (i.e., apply to at least one firm). This seemingly low participation rate is a product of several factors including outdated email addresses; that many business school alum are satisfied and settled with their current career trajectory; and that many people are not interested in working for a startup.[17] For classes graduating in 1980 or before, the participation rate is 0.46%. The participation rate increases with the class year of graduation, and is highest for the class of 2019, where the participation rate is 4.3%.

Appendix Table A1 shows a linear probability model where participation in the RCT is regressed on potential applicant covariates we can observe ex ante on the basis of data from our partner business school alumni lists. Beyond graduation year, men are 0.9pp more likely (i.e., twice as likely) to participate than women. Subjects living in the city where the SEP is headquartered are 0.7pp more likely to participate. The treatment that subjects are assigned does not predict participation, which is unsurprising given that subjects receive identical emails and experiences across treatments until the subject accesses the website.

Firm selection into the RCT. As noted above, 14% of contacted firms agreed to participate. This relatively low participation rate reflects that many startups are not looking to hire at any one time. Startups did not select into the RCT unless they were actively considering job candidates. Appendix Table A2 compares means of firm characteristics of RCT firms and non-participating firms (columns 1-5), whereas column 6 shows a linear probability model of participation in the RCT on firm characteristics. Observable characteristics

[17] We do not view the seemingly low participation rate as a problem. Rather, in a broad set of businesspeople, our study reaches the subset currently interested in applying to startups. Also, our email lists are high-quality and not deficient. Alumni email lists often have outdated emails, especially for older alum.

are generally weak predictors of whether a startup participates in the RCT, suggesting that participating startups are broadly representative of SEP firms.[18]

3 Results

3.1 Impacts on Job applications

Table 3 shows that expert ratings created large shifts in application behavior toward better-quality firms. Figure 4a shows our primary results graphically, plotting estimated treatment effects of business and science information across our four pre-registered measures of application behavior (top-ranked application, application ranked in top 3, any application, normalized rank). For brevity, we will focus on the results in column 3 of Table 3, the specification that simultaneously analyzes both business and science rating treatments.

Starting in Panel A of Table 3, informing applicants that a firm has below-average science decreases the chance that a candidate applies by 7pp or 24% on a base application probability of 28.6%. For firms with above-average science, showing this information increases the chance a given worker applies by 4pp, though this increase is only marginally statistically significant ($p = 0.08$). Candidates are 10pp more likely to apply to a firm that received a positive science rating compared to a negative science rating. Showing that a firm has an above-average business model led to an 8.3pp, or 29%, increase in the probability a worker applies, and showing negative business-model information lowers application likelihood by 3.4pp, or 12%.

In Panel B, negative science information decreases the chance an applicant considers a startup their top choice by 1.2pp (or more than 32% on a base rate of 3.8%), whereas receiving positive information increases it by roughly an equal amount. We receive roughly symmetric results with respect to business model information. Likewise, in Panel C, getting negative science information lowers the chance a startup is ranked in the top 3 by a given applicant by 4.2pp, or 38%, with nearly symmetric effects from viewing positive information. Again, business model information also has roughly symmetric effects in the positive and negative directions, though to a somewhat smaller degree. Finally, in Panel D, receiving negative science information decreases the normalized rank of a given firm by 0.17σ, whereas receiving positive science information increases the normalized rank by 0.12σ. Negative business information decreases the normalized rank by 0.11σ, but receiving positive business information increases normalized rankings by almost 0.16σ.

[18] While not available for all non-participating firms, average science quality is also similar between participating and non-participating firms. Business quality is not available for non-participating firms.

To account for multiple hypothesis testing, Table A7 shows family-wise error rate adjusted p-values based on Westfall & Young (1993), which support our main results reported in Table 3. Overall, we see that job applications respond strongly to expert ratings, both business and science information. Furthermore, effects tend to be roughly symmetric for negative and positive information.

While Table 3 analyzes the decision on applying to particular firms, exploiting randomized within-firm variation, it is also illustrative to look at the overall distribution of firms applied to across treatment arms. This is pursued in Table 5. As seen in column 1, when no expert ratings are shown (i.e., in the control group), firms evaluated to have both above-average science and an above-average business model receive 11% more applications on average than firms graded below-average on each metric (i.e., 21.5% in row 4 vs. 19.3% in row 1). However, as seen in column 4, when workers observe both science and business ratings, startups rated above-average on both metrics receive 80% more applications than those graded below-average.

Magnitudes. How large is the magnitude of our effects on job applications? Belot *et al.* (2018, 2022b) experimentally vary wages of posted jobs in a UK job board and find an elasticity of application-like behavior to posted wage of approximately 0.7. An RCT modifying offered wages in the Mexican Civil Service found an arc-elasticity of approximately 0.8: a 33% increase in offered wages led to 26% more applications (Dal Bo *et al.* (2013)). At those elasticities, our estimated treatment effects from communicating even coarse information that a startup has above-median business model or science quality are able to generate as many additional applications as a 15 to 44% increase in offered wage.

Complements or substitutes? The main analyses in Table 3 are pre-registered and focus on the average effect of providing science and business ratings. One also wonders whether the effects are complementary or not. In Table A8, following Muralidharan *et al.* (2023), we test for complementarity by including an interaction of dummies for receiving business and science ratings, plus this interaction multiplied by whether a firm is good. For clarity, we focus solely on firms that are good in both science and business model, or that are bad in both dimensions. Across all four outcome variables, we consistently see evidence that they are substitutes. A simple explanation for the expert ratings being substitutes is that jobseekers may update favorably about business quality based on positive science ratings, and visa versa, as we show below in Table 4 and Section 3.3. Thus, providing both sets of expert ratings provides less than double the average impact of providing one set of ratings.

3.2 Impacts on Interviews and Hiring

As seen in our RCT registration, we designed our study to be well-powered for examining the impact of our treatments on job applications as they are central for understanding jobseeker preferences. We fully understood that with only 26 participating startups, we would be under-powered to examine later outcomes like interviews and hiring as outcomes.[19] Nonetheless, we conducted a follow-up interview where 19 of 26 startups responded concerning their post-participation interviews and hiring. Of these 19 firms, 4 had interviewed 13 applicants in our sample, and extended 1 formal offer. Five firms had made offers to an early employee, generally either through the founders' network or a technical hire outside the scope of our experiment. These numbers imply call-back rates that are fairly low, but are not atypical at all for high-skill firms like the ones we study.[20]

3.3 Beliefs

Beliefs descriptives. Before showing impacts of information on beliefs, it is first useful to summarize these incentivized beliefs, particularly as they relate to the probability of firm success. Figure 3 summarizes beliefs both on the probability of raising money at a $1m valuation within a year and on the probability of having a successful exit, defined as IPO or acquisition valued at over $50m within a year. There is a lot of heterogeneity, as well as significant bunching at round numbers, consistent with most research using subjective belief data.

Our most striking finding is that respondents dramatically overestimate the probability of a successful exit within one year. While the true probability is essentially zero (i.e., less than 1% of SEP firms have ever had a successful acquisition within a year, and none have had an IPO, figures consistent with seed-stage high-tech startups more broadly), the median answer is 25%. For raising money within a year, the median answer was 52%, compared to the true probability in the data is 23%.

We are not aware of any direct prior evidence that workers substantially overestimate the impact of startup success. These results are particularly noteworthy because (1) the beliefs are strongly incentivized (i.e., people are "putting their money where their mouth is") and (2) the sample includes a large number of experienced workers for whom overestimation is perhaps more surprising than in less-sophisticated samples.

[19] Early-stage startups also tend to hire following financing rounds, and in our follow-up interviews, a number of firms in the primary RCT mentioned that they were still planning to hire once their next tranche of financing was secured. Recall, however, that our primary job board RCT concluded just several months before the beginning of the COVID-19 pandemic.

[20] For example, as of 2019, Google accepted about 0.2% of job candidates. https://www.cnbc.com/2019/04/17/heres-how-many-google-job-interviews-it-takes-to-hire-a-googler.html.

Appendix Figure A2 shows beliefs by worker and firm characteristics. Figure A2a shows that female applicants are particularly likely to overestimate the probability of a raise or IPO, and that the overestimation occurs even among high-quality workers.[21] Panel A of Appendix Table A6 shows the OLS estimates of these results with additional worker characteristics. Interestingly, former startup employees are significantly less optimistic than others about the probability of startup success. A likely explanation is that prior startup experience calibrates the expectations of startup exit.[22]

It is important to note that overconfidence about positive startup outcomes occurs widely across worker characteristics. One concern is that our results could be driven solely by the beliefs of particularly naive or inexperienced applicants. However, as noted, overoptimism is highly prevalent among both high-quality workers, and also among workers with a STEM degree. Note also that because the true probability of an IPO or acquisition within a year for a seed-stage venture is essentially zero, any elicitation or rounding error would lead to overestimation. Moreover, the finding of overprediction is highly robust to excluding multiples of 5 above zero. We also observe that beliefs about firm success are correlated with application decisions (Appendix Table A12).

Impacts of expert ratings on beliefs. Table 4 shows that expert ratings also substantially affect worker beliefs, particularly about perceptions of the business and science quality of firms, but also about firms' perceived chances of raising money and (more tentatively) having a successful exit.

Panel A shows results for normalized perceived science quality. Receiving negative science information lowers perceived science quality by roughly 0.3σ, whereas positive information increases it by roughly 0.15σ. Interestingly, business model ratings also cause individuals to change their beliefs about science quality, but not to the same degree; this is consistent with a prior that assumes correlation between the quality of different aspects of a startup. Panel B shows results for normalized perceived business quality. We see the same qualitative pattern as in Panel A, with relatively strong effects of business information on perceived business quality, and weaker (and here insignificant) effects on perceived science quality but in the expected direction.

Panel C shows that the treatments had substantial effects on workers' beliefs that firms will raise venture capital. Negative science information decreases the perceived chance of a firm raising money by 4pp, very similar to the impact of negative business information. We

[21] While research often finds that men are more overconfident than women, there are also many situations where men and women appear equally overconfident (e.g., Huffman *et al.*, 2022).

[22] We also highlight that we did not ask workers for their beliefs about the firms they applied to, but rather about a fixed, pre-chosen subset of firms on the job board.

also see statistically significant effects of positive business information. Panel D shows that positive business information significantly increases the perceived chance of firms having a major exit. The other coefficients tend to be noisy, likely reflecting the strong heterogeneity in beliefs across workers.

Figure 4b shows these results graphically, plotting estimated treatment effects of business and science information on both perceived firm quality and chances of positive longer-term outcomes.

3.4 Treatment Effect Heterogeneity

This section considers heterogeneity analyses on the extent to which applications respond to expert ratings. We examine heterogeneity according to worker and firm characteristics. There is limited heterogeneity with respect to most characteristics. However, men respond more to science expert ratings than women, and this finding is robust to multiple hypothesis testing correction. Importantly, we find no evidence that low-quality workers—as evaluated by an HR expert focused on startup hiring—are driving our main results.

We address heterogeneity using two methods. First, we examine simple interaction effects in OLS regressions. To maximize statistical power, instead of looking separately at heterogeneity between good and bad ratings, we examine heterogeneity according to overall worker responsiveness to expert ratings. We define the variable BizInfoShock_n (SciInfoShock_n), which is -1 if negative business (science) expert rating is shown, 1 if positive business (science) expert rating is shown, and zero if no information is shown. For each worker characteristic C_n , we estimate a model of the below form:

$$\begin{aligned} y_{nf} = \alpha_0 &+ \alpha_1 \text{BizInfoShock}_n + \alpha_2 \text{SciInfoShock}_n + \alpha_3 \text{C}_n \\ &+ \alpha_4 (\text{BizInfoShock}_n \times \text{C}_n) + \alpha_5 (\text{SciInfoShock}_n \times \text{C}_n) \\ &+ \boldsymbol{X}_{nf} + \varepsilon_{nf}. \end{aligned}$$

For firms, the estimation equation is the same except C_n are firm characteristics.

Second, we apply the sorted effects method of Chernozhukov *et al.* (2018), which uses machine learning to characterize the observations most and least affected by our treatments, and addresses issues of multiple hypothesis testing. Both methods yield the same conclusions.

Measuring worker quality. Worker quality is a key variable in models of sorting (Eeckhout & Kircher, 2011), but is not easily observed. To measure it, we contracted with an independent human resources consultant who specializes in startup hiring, having over 15 years of experience, and asked her to rate workers in terms of their quality on a scale from

1-10. Specifically, she was asked to evaluate worker resumes in terms of their suitability for a management job at a high-tech startup.

Heterogeneity by worker characteristics. Panel (a) Figure 5 examines heterogeneity by worker quality. There is no evidence that lower quality workers respond more to expert ratings; in fact, across all 8 specifications, higher quality workers show greater responsiveness. However, the difference is not statistically significant. Our finding is important because if the marginal worker who switches their application to better firms were low-quality, then information could result in adverse selection, and hence misallocation of human capital.

Panels (b) and (d) of Figure 5 show no difference in responsiveness by whether workers have a STEM degree or by current employment status. Regarding having a STEM degree, on one hand, one might imagine that workers with STEM degrees would have better knowledge about which startups have good science, so that expert ratings are less needed. On the other hand, it is possible that STEM workers intrinsically value good science more than non-STEM workers as an amenity (Stern, 2004), so the two effects could offset each other.

Panel (c) shows that men respond more to information than women, which is robust to adjusting p-values for testing multiple dimension of worker heterogeneity, as shown in Table A9. There are various explanations for this result consistent with prior literature on job search, including that women may be more sensitive to commuting and relocation costs than men (Le Barbanchon *et al.*, 2021). For example, a woman might be interested in applying to a firm rated highly by an expert, but be constrained to do so because the job is far away, whereas a man may be less constrained in relocating (Benson, 2014).[23]

As seen in Appendix Table A10, machine-learning based sorted effects models yield the same qualitative conclusions as those in Figure 5.

Firm characteristics. Figure 6 shows limited evidence of heterogeneity with respect to firm characteristics. There is evidence for a broad picture where workers rely more on signals where there is less information from the firm on that dimension. Using machine learning sorted effects, Appendix Table A10 Panel B shows similar findings to Figure 6.

[23]More broadly, it could be that women value non-pecuniary aspects of jobs more than men, and that the expert rating treatments primarily provide information about the pecuniary return to working at a firm instead of the non-pecuniary return. Separately, one may wonder how our results can be squared with the finding that women are more risk-averse than men (Bertrand, 2011). We observe that all the jobs at the firm in our sample are high-risk (e.g., in terms of possibly losing one's job) even conditional on receiving positive expert ratings, so one would not necessarily expect to respond more to ratings even if they are more risk-averse. Finally, that men respond more to information is not driven by having more statistical power for men (e.g., there being more male than female subjects), as such considerations would affect our ability to detect significant interaction terms. As seen in panel (c) of Figure 5, all 8 point estimates of effects for men are larger than those for women.

Panel (a) of Figure 6 examines heterogeneity based on whether founders have prior business development experience, such as serving as an executive at another company. One might worry less about business model quality for a founder with such experience compared to, say, a computer science professor who has never worked in the private sector (Shaw & Sørensen, 2019). Indeed, for two of our outcomes, we see that workers respond more to expert ratings on business quality when founders do not have prior business development experience.

Panel (b) examines heterogeneity by whether the founder has a PhD, which likely serves as a signal of science quality (and not of business quality). Here, there is weak evidence that workers respond more to science quality when the founders do not have a PhD, consistent with expert ratings serving as a substitute signal.

Panels (c) and (d) examine heterogeneity based on whether a firm is generating positive revenue (many science-based startups take a while to generate revenue), and by whether a firm has received external financing. We do not find consistent evidence that treatment effects depend on these factors.

4 Discussion: Alternative Interpretations, Economist Survey, and Worker Earnings Implications

4.1 Threats to Validity

Hawthorne effects. A common concern in RCTs is whether results are driven by experimental demand effects. However, participants in our study did not know that there was experimental manipulation of what they observe, and firms were only told that the SEP was testing aspects of the job board and hence that there could be experiments performed on its structure both for research and internal-to-SEP purposes. Although workers were told their data may be used for research, it is unlikely that solely knowing this would drive our results: job applications are a relatively high-stakes decision, and hence workers would need to prioritize pleasing researchers in some way over choosing which jobs they were most interested in.

Salience effects. A separate concern is whether simply making any information salient could drive our results (Li & Camerer, 2022). A key feature of our experimental design is that our treatment has equal salience for every firm. Each applicant sees each job marked in a similar fashion (e.g., "Business Model Was Rated Above Average: Yes" and "Business Model Was Rated Above Average: No"), so it is not the case that some jobs receive greater visual

cues than others. In addition, workers behave in ways that suggest response to the particular type of information we are providing and not simply salience effects. For example, providing business ratings affects perceived business quality, but has limited effect on perceived science quality.

A related but different salience concern is whether the treatment could serve as a cheap marker for low-effort jobseekers to make decisions. Perhaps jobseekers are searching for startup jobs across multiple platforms, and the expert ratings provide a quick way of making decisions on the SEP job platform for low-effort jobseekers. We believe that this is unlikely to drive our results given that jobseekers spent substantial time completing the RCTs, both primary (where people spent a median of 22 minutes after clicking on the part of the job board website where they enter information) and secondary RCTs. All our main results are robust to excluding jobseekers who took less than 10 minutes after clicking on the data entry portion during the primary RCT.

Lack of comprehension for the RSD and quadratic scoring rule. Our RCT uses random serial dictatorship (RSD) to allocate applications, and uses a quadratic scoring rule to elicit beliefs. What if workers don't understand these mechanisms or are inattentive? Following past research, we explained the mechanisms in a simple and intuitive manner, emphasizing that people would do best for themselves if they reported their true preferences and beliefs. To the extent that the RSD created measurement error in job application rankings, this would increase the size of our standard errors. This is not a concern for our overall effects, where we find clear statistical significance, though we acknowledge that this may make it harder to detect heterogeneity.

4.2 Are the Results Obvious? Economist Expert Predictions

In order to evaluate the predictability of our findings, we asked economist experts to predict our findings, following DellaVigna & Pope (2018). In April 2023, we sent the survey to 270 economists randomly selected from those who attended the 2022 NBER Summer Institute in Personnel Economics, Entrepreneurship, or Labor Studies, from which we received 86 responses, reflecting a response rate of 32%. The first question asks whether the respondent was already familiar with the main findings of our study, to which 7 respondents said Yes who were immediately screened out and not asked further questions, leaving 79 responses. Of this set, 30% were full professors, 21% were associate professors, and 26% were assistant professors. The remainder were PhD students/postdocs (15%) and industry economists (8%). We also distributed the survey on the Social Science Prediction Platform, where it received an additional 14 responses, of which 4 were screened out, leaving us with 89 total

responses.

The survey focused exclusively on our primary RCT. After the screener question, the survey described our setting and the expert rating treatments, including showing a screenshot of what job seekers saw. Appendix F provides further details and the exact questions on the economist survey.

Table 6 summarizes the results of the economist survey. Column 2 of Table 6 lists actual RCT results, whereas Column 3 summarizes economist predictions. Most economists predict that expert ratings affect job applications, but most economists fail to predict key qualitative features of the effects. For example, our RCT finds comparable effect sizes between science and business ratings, as well as between positive and negative information. However, fewer than 15% of economists predict this, while the majority, 59% and 72%, respectively, believe that business rating and negative information would yield a "significantly larger" impact.

Turning to rows 4-7, economists underestimate the proportional number of applications received by good firms (rated above-average in terms of both science and business) vis-à-vis bad firms (rated below-average in terms of both science and business), when individuals possess access to quality indicators. For instance, when both ratings are shown, we find that good firms received 80% more applications, while the mean expert projection is only 36%. However, experts are relatively accurate, albeit slightly optimistic, about the ability of workers to ascertain startup quality in the absence of expert ratings.[24]

Recall that our RCT does not find heterogeneity in effects based on having a STEM degree or the level of worker quality (as measured by an HR expert). However, less than 30% of economists predict this. The most popular economist response is to predict strong responses for jobseekers who have STEM degrees and are high-quality. While we find larger treatment effects for men than women, only 14% of economists predict this.

4.3 Why Don't Firms Provide More Quality Signals?

If asymmetric information about startup quality is so severe, why don't high-quality startups give credible quality signals? Empirically, they do not. In their self-written ads, 19 of the 26 SEP firms described technical details, 23 described their commercial product, and 8 mentioned their current or planned business model. However, only 4 of the 26 firms

[24]Rows 5-6 show that science and business ratings increased the ratio of applications to good firms to applications to bad firms by 32% and 73%, respectively. How can this be squared with the fact that row 1 shows that science or business ratings had similar effects (based on Table 3)? This reflects that the row 1-3 findings are based on overall effects using all the data, whereas rows 4-7 focus solely on firms that have both good science and good business or that have bad science and bad business.

gave *any* credible quality signal.[25] Our primary explanation is that the founders do not realize that without credible quality signals, their firm is difficult for applicants to evaluate. Alternatively, even top startups may not have outside credible signals to cite. This is not the case for the 26 firms in our primary RCT: we verified in SEP internal documents that all 26 firms had outside signals as defined above which could have been shown in a job ad.

A third explanation is that our sample is unusual. It is, after all, a small sample of startups that are particularly science-heavy. To investigate this pattern more broadly, we examine the content of ads posted on AngelList's hiring board. We scrape the universe of job ads for full-time positions from companies with 1-10 employees who posted a job over two weeks (n=1017).[26] From the full text of the ad, including company descriptions, we hand code whether it describes the company's product, business model, technical details about the product, and credible outside signals of quality, using the same coding as we used for SEP job ads (for details, see Appendix D.4).

Table A11 summarizes this AngelList data. Only 22.7% of ads contain even one such signal, though 93% describe the product being sold and 24.6% even give a technical description of the company's product. The AngelList sample is about as likely to include credible outside signals of quality as the SEP sample, though less likely to provide pure technical description. Nonetheless, it is striking that more than 3 out of every 4 of these startups do not attempt to differentiate their company on quality. Even restricting to non-technical, business development ads (40.4% of the sample), only 24.3% include credible outside signals. That is, the lack of signals which permit job seekers to sort high-quality startups appears to be a general phenomenon, not an unusual quirk of the SEP.

4.4 Translating Firm Quality into Worker Earnings

The following calculation estimates how much a worker's expected lifetime earnings change if they work for the average post-treatment firm they apply to instead of the average pre-treatment firm. Intuitively, more precise information about firm quality leads workers to apply to better firms, better firms are more likely to have a liquidity event, and liquidity events lead to payoffs for early hires via their equity.

[25] We include any mention of the founder's education, whether the firm is a spinout, whether the company participated in incubators, possession of IP, named buyers/partners, existing sales abroad, a named investor, prominent advisor or government grant, a prize or contest victory, named previous experience by founders in a startup or high-level corporate position, any award or prize given to the founders for related work, any media mention, unnamed investors with previous exits, or specific existing sales traction. One firm mentioned their product is based on research they published in a top scientific journal. A second discussed a unique FAA certification. A third discussed their link to an academic lab and their partnership with two major international firms. A fourth discussed the educational background of the founder.

[26] In particular, these are all ads posted between October 30 and November 13, 2020.

In particular, the benefit to a worker from our information treatment is $\mathbb{E}[\Delta_q] \times \frac{\partial R}{\partial q} \times \mathbb{E}[LS]$ where $\mathbb{E}[\Delta_q]$ is the expected change in the expert-estimated quality of firms applications are made to, $\frac{\partial R}{\partial q}$ the marginal increase in the present value of firm revenue R as quality increases, and $\mathbb{E}[LS]$ the expected labor share of revenue that accrues to an early hire.

Both the science and business rating treatments shift roughly 9.5% of applications from below-median to above-median firms.[27] Assuming below-median firms have in expectation 25th-percentile expert-evaluated quality, and above-median ones 75th-percentile quality, each treatment shifts expert-evaluated quality of the average application by $1.35 \times .095 = .128$ standard deviations. A $.128\sigma$ increase in science quality, as shown in Table 2, predicts a 1.2 percentage point increase in the probability a firm raises a venture round after the SEP. Likewise, a similar calculation using just the (noisier) outcomes of firms in the experiment as discussed in Section 1.1 suggests that a $.128\sigma$ increase in science or business model quality raises the probability the firm raises money or hires 10+ employees within 3 years of the experiment by .8 to 2.8 percentage points. Therefore, for each of our information types, $\mathbb{E}[\Delta_q] \times \frac{\partial R}{\partial q}$ is between .8 and 2.8.

As for $\mathbb{E}[LS]$, the increase in remuneration from working at a better firm, Kerr *et al.* (2013) estimate that being funded increases the chance of a positive successful exit by roughly 10pp. Therefore, each of our information treatments leads the average application to be made to a firm with a .08 to .28 percentage point higher chance of a successful exit. There is not good data on the payoff to early workers of a successful exit, but conditional on being venture-backed, one back-of-the-envelope calculation suggests the equity share of an early employee has an expected value of just under $1,000,000.[28] Each information treatment therefore has an expected value to a given applicant of $800 to $2800, solely in terms of working at a startup that is more likely to have an IPO or be acquired.[29]

Beyond workers, one might also be interested in welfare implications of our treatment. A full welfare analysis depends on the importance of assortative matching between workers and firms (i.e., do better firms benefit more from better workers than worse firms) and the degree of total surplus captured by workers, and is beyond the scope of the paper. However, when assortative matching is important to efficient production, welfare effects may exceed

[27] Science ratings raise the probability of an above-median application by 12% and reduce the probability of below-median by 24%. The share of above-median applications post-treatment is thus $\frac{1.12}{1.12+.76} = .596$. Likewise, the post-treatment share of above-median applications for business quality information is .594.

[28] See https://80000hours.org/2015/10/startup-salaries-and-equity-compensation.

[29] This calculation assesses the value to workers of working at the average post-treatment firm applied to compared to the average pre-treatment firm. This is a clear, policy-relevant way of assessing the impact of the treatments on choice quality, but is different, of course, from the average earnings benefit of the treatments (e.g., since the treatments shift applications to better firms, workers may also face more competition). Besides greater pay, working at a better firm may also provide benefits in terms of skill development, career progression, or job stability. Accounting for such benefits would increase the value of the treatments.

the benefits to individual workers.

5 Conclusion

Workers make substantially different job application decisions when randomly given coarse expert opinions on startup quality, shifting applications to better firms. Workers react to both positive and negative expert ratings, and to both information on science and business model quality. Changes in worker beliefs about startup success appear to be one mechanism for these results, though these beliefs also reveal considerable overestimation about the likelihood of exit events.

These results have important policy implications. When a venture capitalist or government wants to invest in a startup, they generally conduct deep diligence on the firm, including obtaining expert opinions. Our results indicate that workers would make quite different job application decisions in the presence of similar information. More precise information about firm quality, especially startup quality, provides an important benefit to both jobseekers and potentially economic efficiency more broadly. On the more pessimistic side, since workers exhibit highly inaccurate beliefs about the chance that startups will have a liquidity event, additional policies besides expert rating may be useful and important in addressing workers' informational deficits.

Turning to implications of the study for SEP itself, the organization was quite pleased with the results of the RCT. SEP's top executive described the results on the impact of expert ratings as "compelling" and others voiced similar opinions. Unfortunately, the timing of the RCT was not good for quick implementation. The main RCT occurred in late 2019, with preliminary results presented to top SEP staff at the end of January 2020. Once Covid hit in early 2020, SEP shifted its focus to re-organizing the program to operate virtually. However, beginning in 2022, SEP leadership informed us that our experiment led them to begin running an annual job board for graduating ventures globally. Moreover, as of spring 2023, SEP has started providing coaching regarding the importance of quality signals in startup hiring.

Our study focuses on high-skilled workers applying to science-based startups. Science-based startups are a growing sector, particularly given the importance of artificial intelligence, and high-skilled workers are natural to study for this, so we believe that our sample is policy-relevant and independently interesting. That said, it is not clear whether our results on substantial information frictions would also hold for more established firms or for young firms in "conventional" industries (e.g., restaurants, construction, etc.). It would be fascinating for this to be explored in future research. Within our sample, treatment effects on

job applications are generally similar across various dimensions of worker background and experience, suggesting that our effects may also hold among broader populations. One could also speculate that our results would be a lower bound for a broader population if lower-skill workers were less sophisticated in their ability to evaluate different firms.

One question left open by our research is why companies do not generally provide credible signals about their quality to prospective employees. In our job board, firms often failed to provide these signals in their self-written job ads. Is it the case that such credible signals do not exist, or are too costly for startups to generate? Or perhaps good startups may overestimate the ability of potential workers to learn about their firm's quality in the absence of these signals? Given the magnitude of the worker application response to these signals of quality, further work is needed on the source of asymmetric information between good firms and good workers, and what factors can reduce these asymmetries.

References

Abdulkadiroglu, Atila, & Sonmez, Tayfun. 1998. Random Serial Dictatorship. *Econometrica*, **66**(3), 689–701.

Aran, Yifat, & Murciano-Goroff, Raviv. 2021. Equity Illusions. *Available at SSRN*.

Artemov, Georgy, Che, Yeon-Koo, & He, Yinghua. 2017. Strategic 'Mistakes': Implications for Market Design Research. *NBER Working Paper*.

Ashraf, Nava, Bandiera, Oriana, Davenport, Edward, & Lee, Scott S. 2020. Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services. *American Economic Review*, **110**(5), 1355–94.

Bandiera, Oriana, Guiso, Luigi, Prat, Andrea, & Sadun, Raffaella. 2015. Matching firms, managers, and incentives. *Journal of Labor Economics*, **33**(3), 623–681.

Belot, Michèle, Kircher, Philipp, & Muller, Paul. 2018. Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice. *Review of Economic Studies*, **86**(4), 1411–1447.

Belot, Michèle, Kircher, Philipp, & Muller, Paul. 2021. *Eliciting Time Preferences When Income and Consumption Vary: Theory, Validation & Application to Job Search*. Tinbergen Institute Discussion Paper 2021-013/V.

Belot, Michèle, Kircher, Philipp, & Muller, Paul. 2022a. *Do the Long-term Unemployed Benefit from Automated Occupational Advice during Online Job Search?* IZA Discussion Paper 15452.

Belot, Michèle, Kircher, Philipp, & Muller, Paul. 2022b. How Wage Announcements Affect Job Search–A Field Experiment. *American Economic Journal: Macroeconomics*, **14**(4), 1–67.

Benson, Alan. 2014. Rethinking the Two-body Problem: The Segregation of Women into Geographically Dispersed Occupations. *Demography*, **51**(5), 1619–1639.

Benson, Alan, Sojourner, Aaron, & Umyarov, Akhmed. 2020. Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market. *Management Science*, **66**(5), 1802–1825.

Bergman, Nittai K., & Jenter, Dirk. 2007. Employee Sentiment and Stock Option Compensation. *Journal of Financial Economics*, **84**(3), 667–712.

Bergman, Peter, Chetty, Raj, DeLuca, Stefanie, Hendren, Nathaniel, Katz, Lawrence F., & Palmer, Christopher. 2019. *Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice*. NBER Working Paper 26164.

Bernstein, Shai, Mehta, Kunal, Townsend, Richard, & Xu, Ting. 2022. *Do Startups Benefit from Their Investors' Reputation? Evidence from a Randomized Field Experiment*. NBER Working Paper 29847.

Bertrand, Marianne. 2011. New Perspectives on Gender. *Pages 1543–1590 of: Handbook of Labor Economics*, vol. 4. Elsevier.

Bloom, Nicholas, & Van Reenen, John. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics*, **122**(4), 1351–1408.

Bloom, Nicholas, & Van Reenen, John. 2011. Human Resource Management and Productivity. *Handbook of Labor Economics*, **1**, 1697–1767.

Bloom, Nicholas, Eifert, Benn, Mahajan, Aprajit, McKenzie, David, & Roberts, John. 2013. Does Management Matter? Evidence from India. *Quarterly Journal of Economics*, **128**(1), 1–51.

Bloom, Nicholas, Brynjolfsson, Erik, Foster, Lucia, Jarmin, Ron, Patnaik, Megha, Saporta-Eksten, Itay, & Van Reenen, John. 2019. What Drives Differences in Management Practices? *American Economic Review*, **109**(5), 1648–83.

Bruine de Bruin, Wändi, Chin, Alycia, Dominitz, Jeff, & van der Klaauw, Wilbert. 2023. Chapter 1 - Household surveys and probabilistic questions. *Pages 3–31 of:* Bachmann, Rüdiger, Topa, Giorgio, & van der Klaauw, Wilbert (eds), *Handbook of Economic Expectations*. Academic Press.

Carreyrou, John. 2018. *Bad Blood: Secrets and Lies in a Silicon Valley Startup*.

Chen, Shuowen, Chernozhukov, Victor, Fernández-Val, Iván, & Luo, Ye. 2019. SortedEffects: sorted causal effects in R. *arXiv preprint arXiv:1909.00836*.

Chen, Yan, & Sönmez, Tayfun. 2002. Improving Efficiency of On-campus Housing: An Experimental Study. *American Economic Review*, **92**(5), 1669–1686.

Chernozhukov, Victor, Fernández-Val, Iván, & Luo, Ye. 2018. The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages. *Econometrica*, **86**(6), 1911–1938.

Dal Bo, Ernesto, Finan, Frederico, & Rossi, Martin A. 2013. Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *Quarterly Journal of Economics*, **128**(3), 1169–1218.

DellaVigna, Stefano, & Paserman, M. Daniele. 2005. Job Search and Impatience. *Journal of Labor Economics*, **23**(3), 527–588.

DellaVigna, Stefano, & Pope, Devin. 2018. Predicting Experimental Results: Who Knows What? *Journal of Political Economy*, **126**(6), 2410–2456.

DellaVigna, Stefano, Lindner, Attila, Reizer, Balázs, & Schmieder, Johannes F. 2017. Reference-dependent Job Search: Evidence from Hungary. *Quarterly Journal of Economics*, **132**(4), 1969–2018.

Dustmann, Christian, Lindner, Attila, Schönberg, Uta, Umkehrer, Matthias, & Vom Berge, Philipp. 2022. Reallocation effects of the minimum wage. *Quarterly Journal of Economics*, **137**(1), 267–328.

Eeckhout, Jan, & Kircher, Philipp. 2011. Identifying Sorting–In Theory. *Review of Economic Studies*, **78**(3), 872–906.

Fadlon, Itzik, Lyngse, Frederik Plesner, & Nielsen, Torben Heien. 2022. *Causal Effects of Early Career Sorting on Labor and Marriage Market Choices: A Foundation for Gender Disparities and Norms*. Mimeo, UCSD.

Flory, Jeffrey A., Leibbrandt, Andreas, & List, John A. 2015. Do Competitive Workplaces Deter Female Workers? A Large-scale Natural Field Experiment on Job Entry Decisions. *Review of Economic Studies*, **82**(1), 122–155.

Gompers, Paul A., Gornall, Will, Kaplan, Steven N., & Strebulaev, Ilya A. 2020. How Do Venture Capitalists Make Decisions? *Journal of Financial Economics*, **135**(1), 169–190.

Hastings, Justine S., & Weinstein, Jeffrey M. 2008. Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *QJE*, **123**(4), 1373–1414.

Hedegaard, Morten Størling, & Tyran, Jean-Robert. 2018. The price of prejudice. *American Economic Journal: Applied Economics*, **10**(1), 40–63.

Hoffman, Mitchell. 2016. How is Information Valued? Evidence from Framed Field Experiments. *The Economic Journal*, **126**(595), 1884–1911.

Hsieh, Chang-Tai, Hurst, Erik, Jones, Charles I, & Klenow, Peter J. 2019. The Allocation of Talent and US Economic Growth. *Econometrica*, **87**(5), 1439–1474.

Hsu, David H., & Ziedonis, Rosemarie H. 2013. Resources as Dual Sources of Advantage: Implications for Valuing Entrepreneurial-firm Patents. *Strategic Management Journal*.

Huffman, David, Raymond, Collin, & Shvets, Julia. 2022. Persistent Overconfidence and Biased Memory: Evidence from Managers. *American Economic Review*, **112**(10), 3141–75.

Ichniowski, Casey, Shaw, Kathryn, & Prennushi, Giovanna. 1997. The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *American Economic Review*, **87**(3), 291–313.

Jäger, Simon, Roth, Christopher, Roussille, Nina, & Schoefer, Benjamin. 2022. *Worker Beliefs about Outside Options*. NBER Working Paper 29623.

Kerr, William, Lerner, Josh, & Schoar, Antoinette. 2013. The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. *Review of Financial Studies*.

Kling, Jeffrey R., Mullainathan, Sendhil, Shafir, Eldar, Vermeulen, Lee C., & Wrobel, Marian V. 2012. Comparison Friction: Experimental Evidence from Medicare Drug Plans. *Quarterly Journal of Economics*, **127**(1), 199–235.

Le Barbanchon, Thomas, Rathelot, Roland, & Roulet, Alexandra. 2021. Gender Differences in Job Search: Trading Off Commute Against Wage. *QJE*, **136**(1), 381–426.

Lerner, Josh. 1995. Venture Capitalists and the Oversight of Privately-Held Firms. *Journal of Finance*, **50**(1), 301–318.

Li, Xiaomin, & Camerer, Colin F. 2022. Predictable Effects of Visual Salience in Experimental Decisions and Games. *Quarterly Journal of Economics*, **137**(3), 1849–1900.

Manchester, Colleen, Benson, Alan, & Shaver, J. Myles. 2023. Dual Careers and the Willingness to Consider Employment in Startup Ventures. *Strategic Management Journal*, Forthcoming.

McKelvey, Richard, & Page, Talbot. 1990. Public and Private Information: An Experimental Study of Information Pooling. *Econometrica*, **58**(6), 1321–1339.

Muralidharan, Karthik, Romero, Mauricio, & Wüthrich, Kaspar. 2023. Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments. *Review of Economics and Statistics*, Forthcoming.

Oyer, Paul. 2004. Why Do Firms Use Incentives That Have No Incentive Effects? *Journal of Finance*, **59**(4), 1619–1650.

Oyer, Paul, & Schaefer, Scott. 2005. Why Do Some Firms Give Stock Options to All Employees?: An Empirical Examination of Alternative Theories. *Journal of Financial Economics*, **76**(1), 99–133.

Oyer, Paul, & Schaefer, Scott. 2011. Personnel Economics: Hiring and Incentives. *Handbook of Labor Economics*.

Roberts, John, & Shaw, Kathryn. 2022. *Managers and the Management of Organizations*. Working Paper 30730. National Bureau of Economic Research.

Shaw, Kathryn, & Sørensen, Anders. 2019. The Productivity Advantage of Serial Entrepreneurs. *ILR Review*, **72**(5), 1225–1261.

Sorenson, Olav, Dahl, Michael, Canales, Rodrigo, & Burton, Diane. 2021. Do Startup Employees Earn More in the Long Run? *Organization Science*, **32**(3), 587–604.

Spinnewijn, Johannes. 2015. Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association*, **13**(1), 130–167.

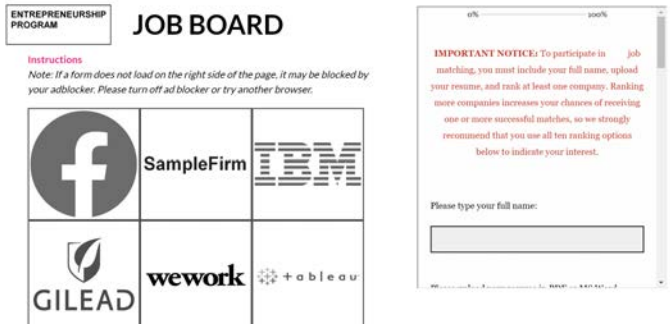
Stern, Scott. 2004. Do Scientists Pay to be Scientists? *Management Science*, **50**(6), 835–853.

Syverson, Chad. 2011. What Determines Productivity? *Journal of Economic Literature*, **49**(2), 326–65.

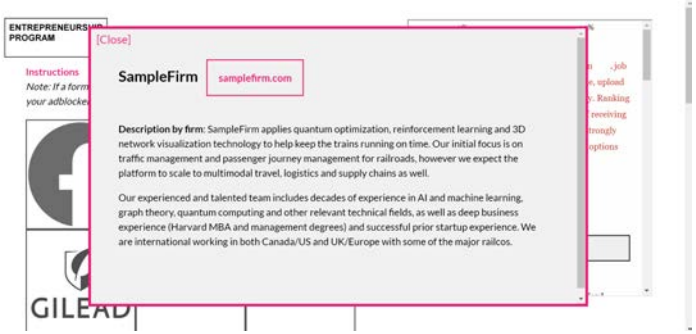
Westfall, Peter H, & Young, S Stanley. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279. John Wiley & Sons.

Wiswall, Matthew, & Zafar, Basit. 2015. Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies*, **82**(2), 791–824.

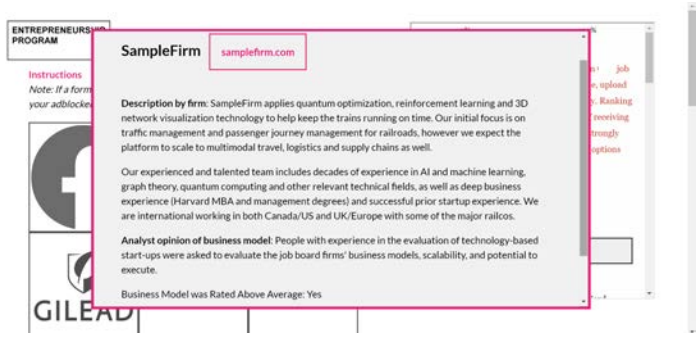
Figure 1: Screenshots from the RCT Job Board



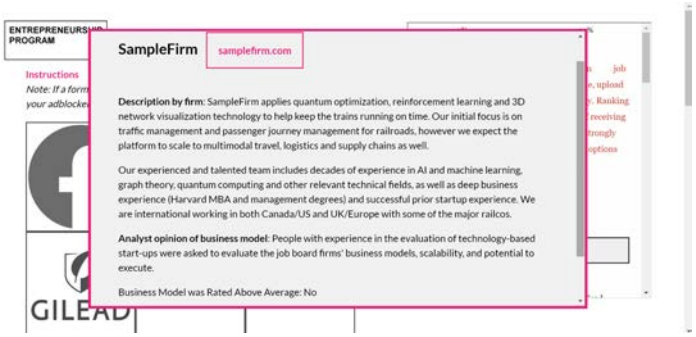
(a) Job board



(b) Control



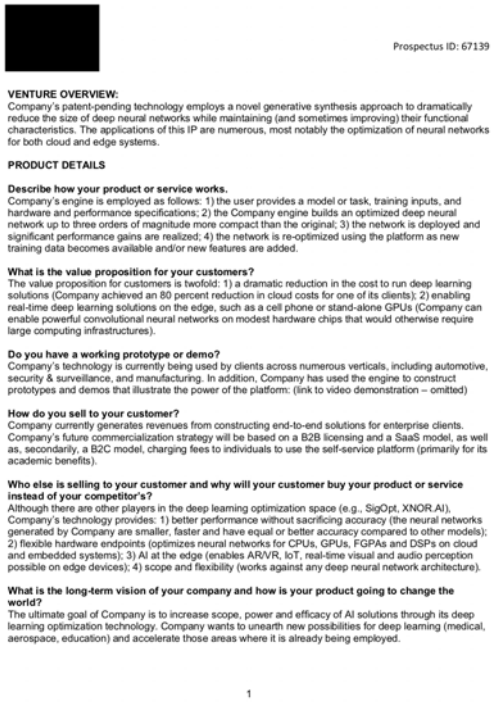
(c) Business rating only, positive



(d) Business rating only, negative

Notes: This figure presents screenshots from the RCT job board. Identifying information about the SEP has been redacted. The description of SampleFirm is based on a mixture of actual firm-written ads. To preserve anonymity of the job board, the logos of actual firms have been replaced with the logos of well-known startups. All of the actual startups in the RCT were between seed stage and Series A.

Figure 2: Screenshots for the Secondary RCT



[redacted]

Prospectus ID: 67139

VENTURE OVERVIEW:
Company's patent-pending technology employs a novel generative synthesis approach to dramatically reduce the size of deep neural networks while maintaining (and sometimes improving) their functional characteristics. The applications of this IP are numerous, most notably the optimization of neural networks for both cloud and edge systems.

PRODUCT DETAILS

Describe how your product or service works.
Company's engine is employed as follows: 1) the user provides a model or task, training inputs, and hardware and performance specifications; 2) the Company engine builds an optimized deep neural network up to three orders of magnitude more compact than the original; 3) the network is deployed and significant performance gains are realized; 4) the network is re-optimized using the platform as new training data becomes available and/or new features are added.

What is the value proposition for your customers?
The value proposition for customers is twofold: 1) a dramatic reduction in the cost to run deep learning solutions (Company achieved an 80 percent reduction in cloud costs for one of its clients); 2) enabling real-time deep learning solutions on the edge, such as a cell phone or stand-alone GPUs (Company can enable powerful convolutional neural networks on modest hardware chips that would otherwise require large computing infrastructures).

Do you have a working prototype or demo?
Company's technology is currently being used by clients across numerous verticals, including automotive, security & surveillance, and manufacturing. In addition, Company has used the engine to construct prototypes and demos that illustrate the power of the platform: (link to video demonstration – omitted)

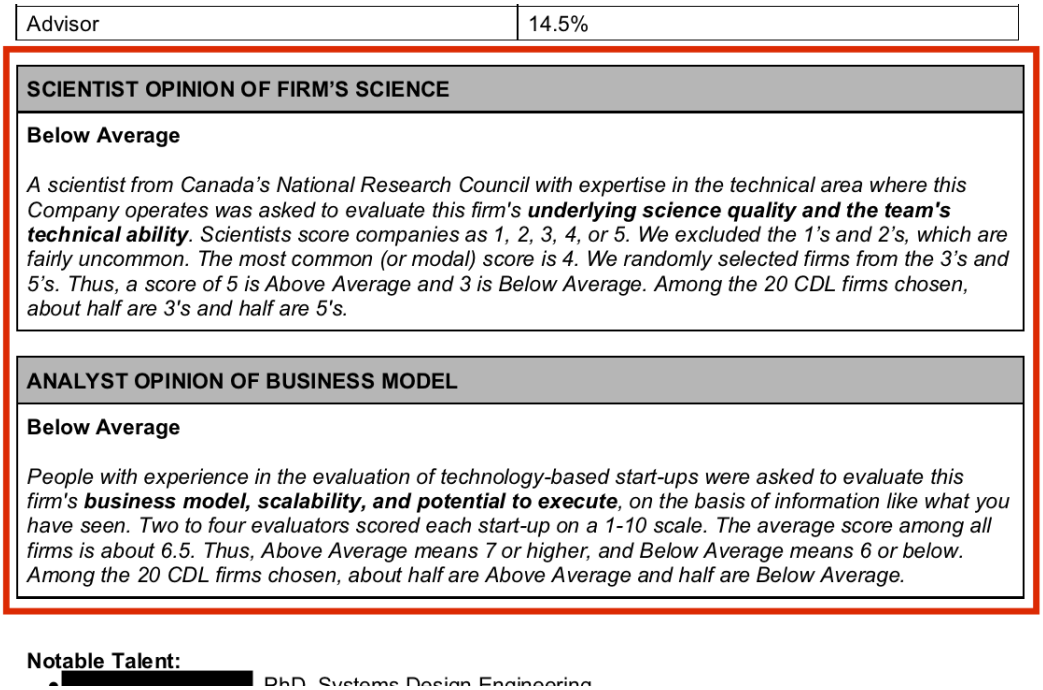
How do you sell to your customer?
Company currently generates revenues from constructing end-to-end solutions for enterprise clients. Company's future commercialization strategy will be based on a B2B licensing and a SaaS model, as well as, secondarily, a B2C model, charging fees to individuals to use the self-service platform (primarily for its academic benefits).

Who else is selling to your customer and why will your customer buy your product or service instead of your competitor's?
Although there are other players in the deep learning optimization space (e.g., SigOpt, XNOR.AI), Company's technology provides: 1) better performance without sacrificing accuracy (the neural networks generated by Company are smaller, faster and have equal or better accuracy compared to other models); 2) flexible hardware endpoints (optimizes neural networks for CPUs, GPUs, FGPAs and DSPs on cloud and embedded systems); 3) AI at the edge (enables AR/VR, IoT, real-time visual and audio perception possible on edge devices); 4) scope and flexibility (works against any deep neural network architecture).

What is the long-term vision of your company and how is your product going to change the world?
The ultimate goal of Company is to increase scope, power and efficacy of AI solutions through its deep learning optimization technology. Company wants to unearth new possibilities for deep learning (medical, aerospace, education) and accelerate those areas where it is already being employed.

1

(a) Control group dossier



Advisor	14.5%

SCIENTIST OPINION OF FIRM'S SCIENCE

Below Average

A scientist from Canada's National Research Council with expertise in the technical area where this Company operates was asked to evaluate this firm's ***underlying science quality and the team's technical ability****. Scientists score companies as 1, 2, 3, 4, or 5. We excluded the 1's and 2's, which are fairly uncommon. The most common (or modal) score is 4. We randomly selected firms from the 3's and 5's. Thus, a score of 5 is Above Average and 3 is Below Average. Among the 20 CDL firms chosen, about half are 3's and half are 5's.*

ANALYST OPINION OF BUSINESS MODEL

Below Average

People with experience in the evaluation of technology-based start-ups were asked to evaluate this firm's ***business model, scalability, and potential to execute****, on the basis of information like what you have seen. Two to four evaluators scored each start-up on a 1-10 scale. The average score among all firms is about 6.5. Thus, Above Average means 7 or higher, and Below Average means 6 or below. Among the 20 CDL firms chosen, about half are Above Average and half are Below Average.*

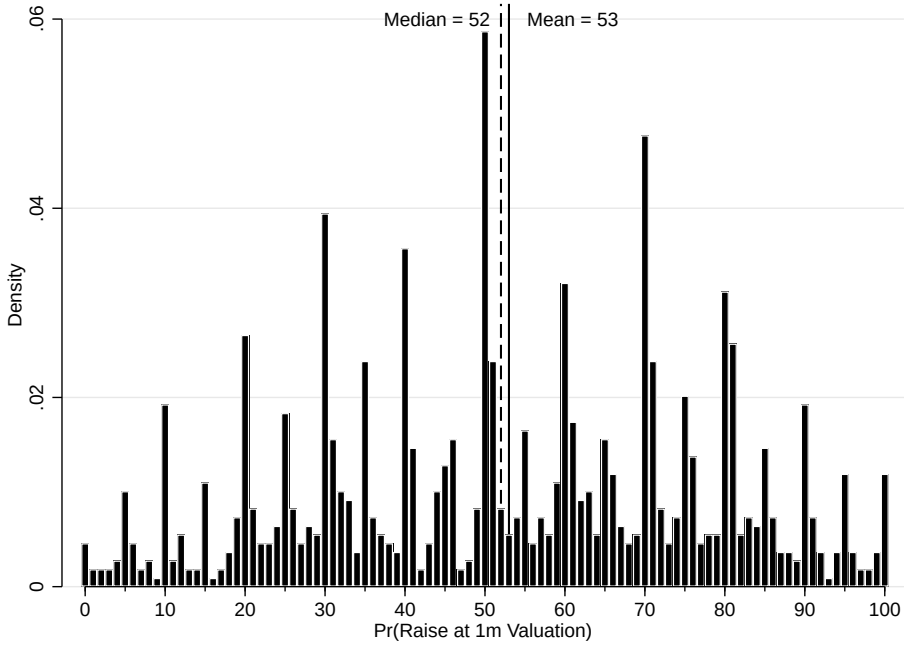
Notable Talent:

[redacted] PhD, Systems Design Engineering

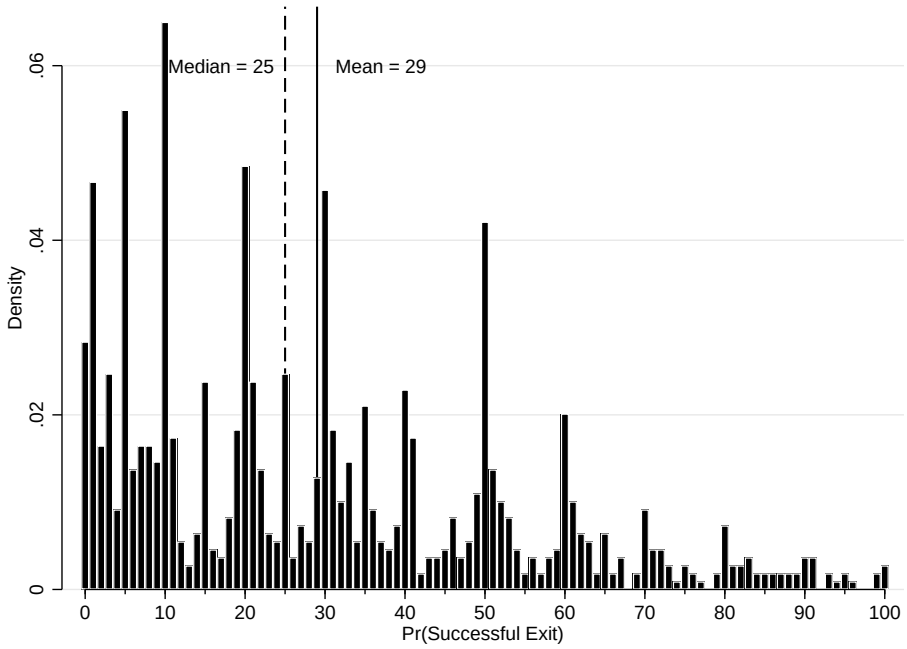
(b) Additional Treatment Group Info

Notes: This figure shows screenshots from the Secondary RCT. Panel (a) shows a sample page from a startup dossier MBA students received to evaluate firms. Panel (b) shows an example of the information provided to the treatment group that received both business model and science scores.

Figure 3: Histograms of Worker Beliefs



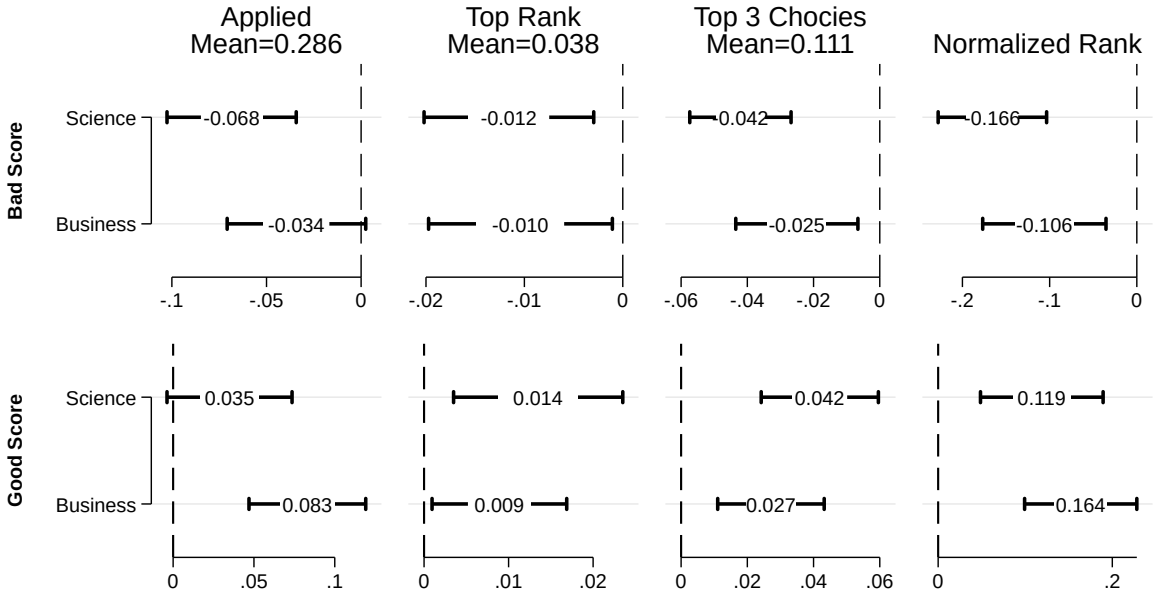
(a) Raise at $1m Valuation in 1 Year



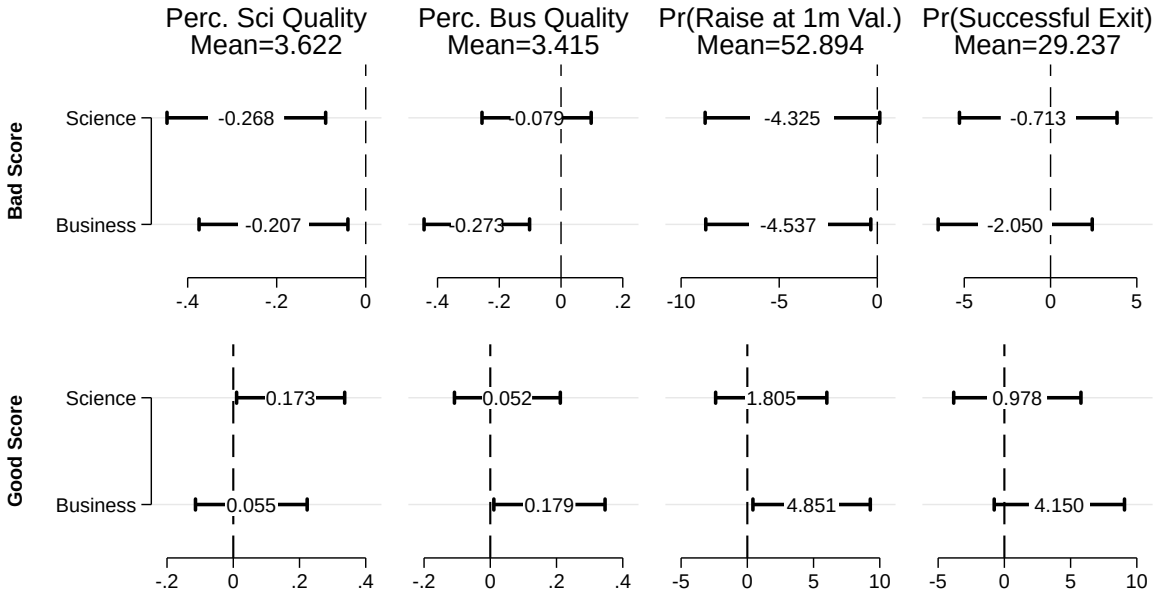
(b) Major Exit (IPO or $50m Acquisition) in 1 Year

Notes: This figure shows histograms of worker beliefs about firm success elicited using a risk-invariant quadratic scoring rule. Data is pooled beliefs from the Primary and Secondary RCTs.

Figure 4: Visual Summary of the Effect of Expert Ratings on Job Applications and Beliefs



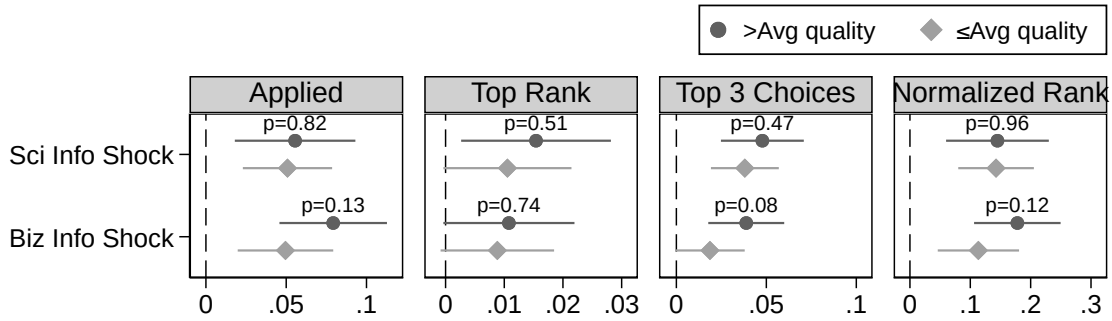
(a) Effect of showing positive and negative expert ratings on job application rankings



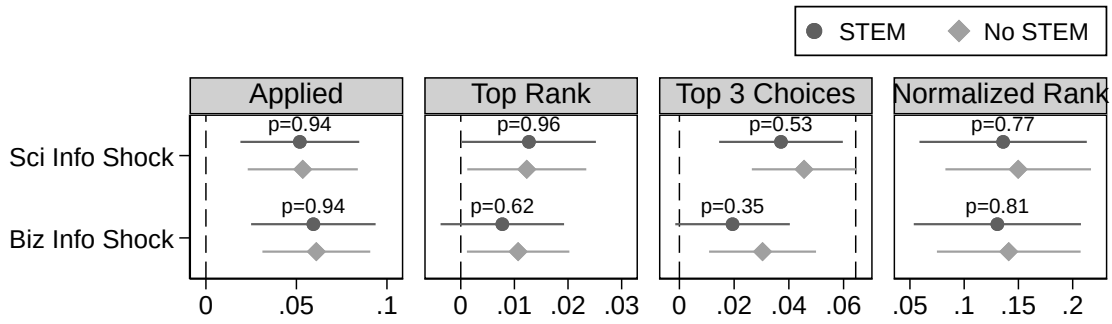
(b) Effect of positive and negative expert ratings on worker beliefs

Notes: Figure 4a plots the estimated effect of expert ratings on job applications from the Primary RCT. Figure 4b shows the same effect on worker beliefs using pooled data from the Primary and Secondary RCTs. *Bad/Good Score* estimates show percentage point change in the probability of applying to a bad/good firm conditional on showing expert ratings. The left two subgraphs show standard error change in the perceived quality of the firm's science and business. The right two subgraphs show the percentage point change in the estimated probability of raising capital at a valuation of at least $1 million within 1 year and successful exit via IPO or getting acquired for at least $50 million within 1 year. All confidence intervals depicted are at 95% level.

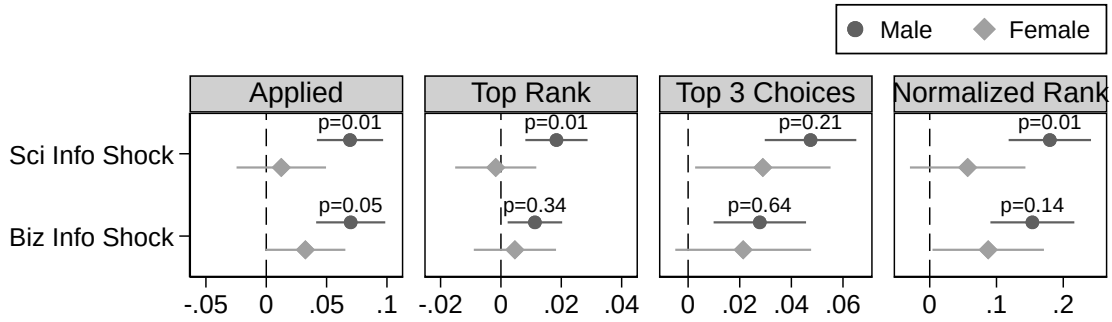
Figure 5: Treatment Effect Heterogeneity for Job Applications by Worker Characteristics



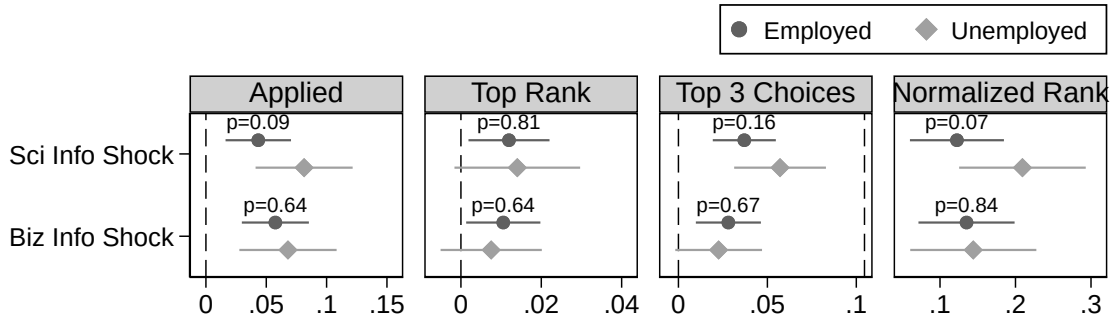
(a) By worker quality



(b) By whether the worker has a STEM undergraduate degree



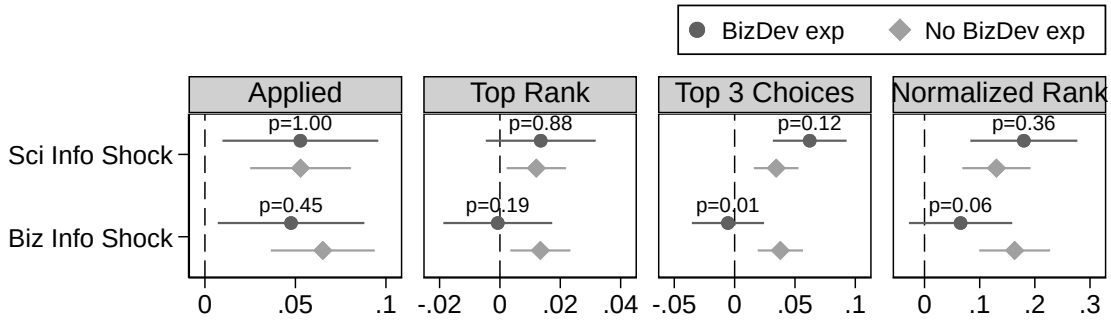
(c) By gender



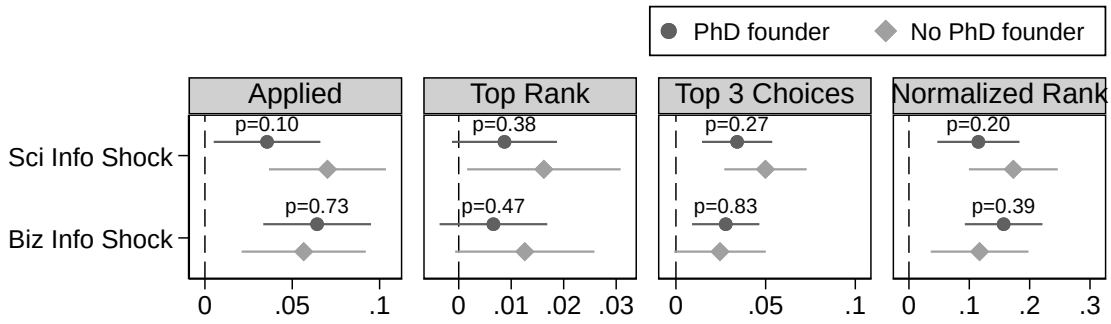
(d) By current employment status

Notes: This figure shows heterogeneity in worker response to science and business information shocks in the Primary RCT. Estimates are from regressing job application outcomes on science and business treatments and their interactions with worker characteristics. *Sci Info Shock* (*Bus Info Shock*) is -1 if negative science (business) information is shown, 1 if positive science (business) information is shown, and zero if no information is shown. Regressions include venture and strata fixed effects.

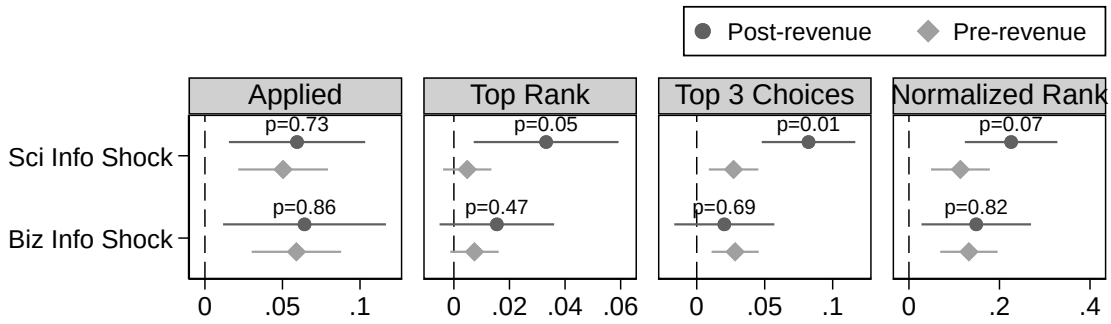
Figure 6: Treatment Effect Heterogeneity for Job Applications by Firm Characteristics



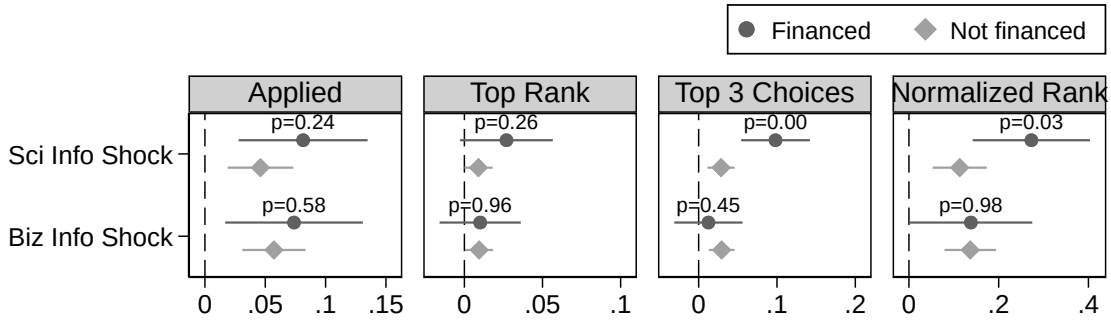
(a) By whether founders have prior business development experience



(b) By whether the firm has a PhD founder



(c) By whether the firm is generating revenue



(d) By whether the firm is externally financed

Notes: This figure shows heterogeneity in worker response to science and business information shocks in the Primary RCT. Estimates are from regressing job application outcomes on science and business treatments and their interactions with worker characteristics. *Sci Info Shock* (*Bus Info Shock*) is -1 if negative science (business) information is shown, 1 if positive science (business) information is shown, and zero if no information is shown. Regressions include venture and strata fixed effects.

Table 1: Balance Table

	No Info	Science Info	Business Info	Science & Business Info	p-value
Panel A: Job Board RCT					
Male	0.77	0.72	0.79	0.69	0.51
City is SEP HQ	0.56	0.47	0.52	0.47	0.70
Graduation year	2012	2012	2013	2012	0.69
Startup founder	0.24	0.23	0.12	0.12	0.14
Startup employee	0.27	0.28	0.19	0.28	0.60
Employed	0.80	0.81	0.69	0.70	0.24
Yrs of exp	9.95	10.69	8.64	10.36	0.55
STEM	0.38	0.49	0.39	0.36	0.49
Worker Quality (1-10)	5.20	5.66	4.90	5.00	0.35
Num. Workers	66	53	67	64	
Panel B: MBA Student RCT					
Male	0.54	0.38	0.41	0.44	0.39
White	0.26	0.29	0.24	0.21	0.82
Hisp./Latino	0.09	0.08	0.06	0.06	0.94
Asian	0.30	0.33	0.49	0.35	0.25
Num. Workers	46	48	49	48	

Notes: This table compares applicant characteristics across treatment groups for the Primary RCT (Panel A) and Secondary RCT (Panel B). Randomization is stratified on gender, city, and year of graduation at the time potential applicants were contacted. Only the Primary RCT's randomization was stratified based on gender, whether the worker lives in the same city as SEP headquarters, and year of graduation. Other variables were not observable prior to application. "Science info" and "Business Info" in column headers refer to the subjects that received science and business scores. Worker Quality is based on a startup-focused HR expert's evaluation of resume quality. Of 259 resumes submitted in the job board, 9 were ineligible and were removed from all analysis. Ineligible candidates were forwarded the link to the job board from eligible candidates. The remaining 19,109 individuals contacted did not apply to any firm.

Table 2: The Correlation Between Expert Science Ratings and Firm Outcomes

	(1)	(2)
Panel A: Dep. Var. = Graduated from SEP		
Science quality	0.102*	0.092*
	(0.055)	(0.050)
R^2	0.03	0.10
Observations	106	106
Mean of DV		0.41
Panel B: Dep.Var. = Raised After SEP		
Science quality	0.087*	0.094**
	(0.048)	(0.045)
R^2	0.03	0.04
Observations	106	106
Mean of DV		0.25

Notes: This table shows results from regression start-up success outcomes roughly four years after participating in SEP on expert science scores given at the time the startup applied to SEP. Data is the full cohort of 130 startups in 2017-18, from which 24 startups were dropped due to missing science scores. Robust standard errors in parentheses. Column 1 has no controls while column 2 controls for the number of founders, an indicator for firms with a PhD founder, and technology fixed effects.

Table 3: The Effect of Expert Ratings on Job Applications

	(1)	(2)	(3)
Panel A: Dep. Var. = Applied			
Science info X Good science	0.102***		0.103***
	(0.025)		(0.025)
Science info	−0.067***		−0.068***
	(0.018)		(0.017)
Business info X Good business		0.116***	0.117***
		(0.026)	(0.026)
Business info		−0.034*	−0.034*
		(0.019)	(0.019)
F(Sci + Sci X GoodSci = 0)	0.076		0.078
F(Bus + Bus X GoodBus = 0)		0.000	0.000
Mean of DV			0.286
Panel B: Dep. Var. = Top Rank			
Science info X Good science	0.025***		0.025***
	(0.009)		(0.009)
Science info	−0.011***		−0.012***
	(0.004)		(0.004)
Business info X Good business		0.019**	0.019**
		(0.009)	(0.009)
Business info		−0.010**	−0.010**
		(0.005)	(0.005)
F(Sci + Sci X GoodSci = 0)	0.009		0.009
F(Bus + Bus X GoodBus = 0)		0.032	0.029
Mean of DV			0.038
Panel C: Dep. Var. = Top 3 Choices			
Science info X Good science	0.083***		0.084***
	(0.016)		(0.016)
Science info	−0.042***		−0.042***
	(0.008)		(0.008)
Business info X Good business		0.051***	0.052***
		(0.017)	(0.017)
Business info		−0.025***	−0.025***
		(0.009)	(0.009)
F(Sci + Sci X GoodSci = 0)	0.000		0.000
F(Bus + Bus X GoodBus = 0)		0.001	0.001
Mean of DV			0.111
Panel D: Dep. Var. = Normalized Rank			
Science info X Good science	0.282***		0.285***
	(0.057)		(0.057)
Science info	−0.163***		−0.166***
	(0.032)		(0.032)
Business info X Good business		0.267***	0.270***
		(0.059)	(0.058)
Business info		−0.106***	−0.106***
		(0.036)	(0.036)
F(Sci + Sci X GoodSci = 0)	0.001		0.001
F(Bus + Bus X GoodBus = 0)		0.000	0.000
Observations	6,500	6,500	6,500

Notes: This table shows the effect of information on employee ranking of 26 startups in the Primary RCT. Dependent variables in Panels A, B, and C are indicators that equal to 1 if, respectively, the worker applied to the firm, the firm is the worker's top choice, and the firm is in the worker's top 3 choices. The dependent variable in Panel D is the normalized rank of the firms. Regressions include startup fixed effects and randomization strata based on gender, whether the worker lives in the same city as SEP headquarters, and year of graduation. Since the year of graduation strata used in randomization are different across the different schools and programs, we combine them here into 3 bins. Standard errors clustered by worker in parentheses.

Table 4: The Impact of Expert Ratings on Worker Beliefs

	(1)	(2)	(3)
Panel A: Dep. Var. = Perc. Sci Quality			
Science info X Good science	0.444***		0.441***
	(0.115)		(0.115)
Science info	−0.269***		−0.268***
	(0.091)		(0.091)
Business info X Good business		0.264**	0.262**
		(0.111)	(0.112)
Business info		−0.210**	−0.207**
		(0.086)	(0.085)
F(Sci + Sci X GoodSci = 0)	0.035		0.038
F(Bus + Bus X GoodBus = 0)		0.528	0.526
Observations	1,094	1,094	1,094
Panel B: Dep. Var. = Perc. Biz Quality			
Science info X Good science	0.129		0.131
	(0.115)		(0.115)
Science info	−0.075		−0.079
	(0.091)		(0.090)
Business info X Good business		0.451***	0.451***
		(0.115)	(0.116)
Business info		−0.273***	−0.273***
		(0.087)	(0.087)
F(Sci + Sci X GoodSci = 0)	0.513		0.528
F(Bus + Bus X GoodBus = 0)		0.037	0.038
Observations	1,095	1,095	1,095
Panel C: Dep. Var. = Pr(Raise at 1m Valuation)			
Science info X Good science	6.084**		6.130**
	(2.820)		(2.838)
Science info	−4.241*		−4.325*
	(2.259)		(2.264)
Business info X Good business		9.343***	9.388***
		(2.797)	(2.804)
Business info		−4.517**	−4.537**
		(2.138)	(2.139)
F(Sci + Sci X GoodSci = 0)	0.389		0.400
F(Bus + Bus X GoodBus = 0)		0.033	0.032
Observations	1,090	1,090	1,090
Panel D: Dep. Var. = Pr(Successful Exit)			
Science info X Good science	1.664		1.691
	(2.793)		(2.800)
Science info	−0.618		−0.713
	(2.311)		(2.322)
Business info X Good business		6.200**	6.200**
		(2.709)	(2.713)
Business info		−2.036	−2.050
		(2.267)	(2.270)
F(Sci + Sci X GoodSci = 0)	0.671		0.690
F(Bus + Bus X GoodBus = 0)		0.098	0.099
Observations	1,092	1,092	1,092

Notes: This table shows the effect of information on employee beliefs using pooled data from the Primary and Secondary RCTs. In Panel A, the dependent variable is the perceived quality of the firm's science on a scale of 1 - 5 (highest), in Panel B it is the perceived quality of the firm's business on a scale of 1- 5, in Panel C it is the estimated probability of the firm raising capital at a valuation of at least $1 million within 1 year, and in Panel D it is the estimated probability of the firm making an IPO or getting acquired for at least $50 million within 1 year. Standard errors clustered by worker in parentheses.

Table 5: Share of Job Applications to Firms Under Different Treatments: Showing Expert Ratings Shifts Applications to Firms with Better Ratings

	Treatment Group			
Firm Type	Control	Science Info	Business Info	Science & Biz Info
Bad Sci Bad Biz	0.193	0.118	0.145	0.142
Good Sci Bad Biz	0.211	0.361	0.171	0.228
Bad Sci Good Biz	0.381	0.329	0.435	0.374
Good Sci Good Biz	0.215	0.191	0.250	0.256

Notes: This table shows the share of applications to firms under different treatments in the Primary RCT. For example, the second column shows that for jobseekers randomly assigned to see science expert ratings (in addition to the baseline information in the job board), 12% of their applications were made to firms with bad science and bad business ratings, 36% of their applications were made to firms with good science and bad business ratings, 33% were made to firms with bad science and good business ratings, and 19% were made to firms with good science and good business ratings.

Table 6: Economist Expert Survey Results: Findings are not Obvious

Prediction problem	Actual	Expert Prediction
Was the effect larger for:		
1. Science or business rating	Similar effect	Larger for sci 22%, Larger for biz 59% **Similar Effect 15%**, No effect 4%
2. Positive or negative rating	Similar effect	Larger for +ve 14%, Larger for -ve 72% **Similar effect 11%**, No effect 3%
Were science/business ratings:		
3. Complements or substitutes	Substitutes	Complements 67%, **Substitutes 27%**, No effect 6%
What percent more/less apps to good sci & biz firms compared to bad sci & biz firms when:		
4. No ratings shown	11%	Mean = 15 p25=3.5%, p50=10%, p75=20%
5. Science ratings shown	32%	Mean = 23 p25=10%, p50=19%, p75=30%
6. Business rating shown	73%	Mean = 28 p25=10%, p50=20%, p75=40%
7. Both ratings shown	80%	Mean = 36 p25=12%, p50=25%, p75=50%
Did the treatment effect vary by worker:		
8. STEM degree	No difference	Smaller with STEM 30%, **No difference 29%** Larger with STEM 41%, No effect 0%
9. Quality	No difference	Smaller for high-quality 18%, **No difference 26%** Larger for high-quality 53%, No effect 3%
10. Gender	Smaller for women	**Smaller for women 14%**, Larger for women 44% No difference 42%, No effect 0%

Notes: This table summarizes expert predictions of our RCT findings by 89 economists. The first column shows the abbreviated form of the prediction questions in the order shown to responders, the second column shows our results, and the third column shows economist predictions. The percentages shown for nonnumeric questions 1 to 3 and 8 to 10 indicate the fraction of responses for each category. Complete questions and information shown to responders are described in Appendix F.