Information Frictions and Employee Sorting Between Startups*

Kevin A. Bryan† Mitchell Hoffman‡ Amir Sariri§

July 2024

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.

Keywords: Hiring, job applications, startups, overconfidence
JEL Classifications: M50, M51

*We are grateful to Nick Bloom, Christian Dustmann, Florian Englmaier, Claudine Gartenberg, Matt Gentzkow, Jonah Rockoff, Kathryn Shaw, and especially Michèle Belot and Alan Benson for detailed comments. We also thank numerous seminar and conference participants for their comments. 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. 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.

†U. Toronto Rotman School of Management

‡UC Santa Barbara and U. Toronto Rotman School of Management and NBER and CEPR

§Purdue University Daniels School of Business

1 Introduction

Hiring is a central issue in economics and especially in personnel economics (Ichniowski *et al.*, 1997; Staiger & Rockoff, 2010; Oyer & Schaefer, 2011; Bloom & Van Reenen, 2011). As surveyed by Benson & Shaw (2024), 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 was made available to workers, would they change who they apply to? Would they apply to *ex ante* better firms? We address these questions using two 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 startups (described more in Section 2). After these 26 startups provide job ads, we build a custom job board accessible to nearly 20,000 business school alumni. Besides startups' own ads, some 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 settings, applicants are free to use firm websites, press coverage, and so on to investigate firms they may apply to. In total, 1,877 applications are made by 250 job-seekers.

Applicants are unaware that information is randomized. 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 ads). 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* obvious how workers would react to expert ratings of science or business quality. They will not react if they already have precise information about these dimensions of firm quality.[1] They will also not react much if they don't care much about these features. For instance, they may focus on other job characteristics like salary, industry, and city. Expert ratings predict firm success, both in past data and for 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 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 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 ($\approx$ \$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 (i.e., a high-value acquisition or IPO) within a year. Despite this, workers update their beliefs in response

[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 underlying science was poor (Carreyrou, 2018). In our data, startups' science and business quality scores are uncorrelated, as detailed in Section 2.1.

to expert ratings, particularly 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 find limited heterogeneity for most dimensions of worker and firm quality. There is no significant heterogeneity by worker startup experience or STEM background, though 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 results were obvious, we consider this both via an economist survey and via the revealed preference of startups in the RCT. 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 the qualitative finding that expert ratings would affect applications, but substantially underpredict the quantitative magnitude. While we find that high-rated firms get 80% more applications than low-rated firms when both ratings are shown, the median economist prediction is a difference of only 25%, and 86% of economists underpredict the magnitude. In terms of revealed preference, our RCT was structured such that startups wrote their own job ads without any restrictions on our end. Despite this, only 4 of our 26 startups mentioned any hard business or science quality signal, a percentage in line with what we find in AngelList ads more broadly. That is, startups themselves don't appear to find it obvious that applicants are unable to separate the startup wheat from the chaff.

Our paper contributes to four literatures. First, it contributes to work in personnel economics, organizational economics, and labor economics on worker/firm matching, 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

[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); Leibbrandt & List (2018); Wiswall & Zafar (2015). Unlike us, 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 new firms on mTurk and show that randomly endowed better reputations increase job fill rates. Working with a staffing agency that staffs for thousands of firms, Bapna *et al.* (2021) show that rejection email content substantially affects whether workers apply to firms in the future. Appendix B.1 discusses more on related work outside of economics.

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; Bloom *et al.*, 2013, 2019) are appreciated by workers.

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* (Ashraf *et al.*, 2022; Dustmann *et al.*, 2017). Of particular relation to our paper is 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 discussion by legal and business scholars on potential regulation of worker pay at startups (Aran & Murciano-Goroff, 2023).

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 (Belot *et al.*, 2021), overconfidence (Spinnewijn, 2015), and rejection aversion (Bapna

[3] Belot *et al.* (2018, 2022a) study information frictions about what types of jobs are available or would be a fit, while we study information frictions over firm quality. Dustmann *et al.* (2022) show that a non-information policy, higher minimum wages, reallocates workers to better firms. Jäger *et al.* (2024) show workers have biased beliefs about their outside option in the broad German labor market. Non-job RCTs study information frictions in other areas (e.g., Hastings & Weinstein, 2008).

[4] There is much work on CEO overconfidence. Our focus is overconfidence of lower-level workers.

et al., 2021). 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 on resource acquisition. Work examines why good startups have trouble acquiring resources like capital and partnerships (Lerner, 1995; Hsu & Ziedonis, 2013). Our work helps address why good startups have difficulty getting workers, and suggests good startups may have trouble hiring because workers don't know which startups are most promising.

2 Context, RCT Design, & Theoretical Framework

Context. SEP is a non-profit, selective, nine-month program operating across several high-profile business schools in seven countries (as of 2023). The core mission is to provide mentorship to early-stage, science-based startups. SEP is akin to a forum that convenes angel and institutional investors and technical experts five times a year to offer structured mentorship.[5] At the time of our RCTs, about 900 ventures had gone through SEP, including 389 in the 2018-19 cohort, indicating SEP's rapid growth.

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

To support startups with distinct technological trajectories, SEP is offered through specialized streams such as machine learning, quantum machine learning, space, health, and energy. Streams use staff and mentors with corresponding experience or expertise. 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 esteemed programs of its kind in the world. SEP's stature means it accesses world-class science and business expert raters, and also that a sizable number of startups want to participate in the RCT.

Second, SEP startups are the type of startups who frequently engage in hiring. SEP is not 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 thought to possess

[5] SEP's mentorship structure aims to help startups design and prioritize short-term measurable objectives. For brevity, we avoid describing aspects of SEP not relevant to our study.

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

Third, ours is a natural setting for analyzing uncertainty over business and science quality, as SEP firms tackle frontier science and business problems. This is a key feature of SEP given our research questions, though we fully acknowledge that our findings may not necessarily apply to all young firms (e.g., firms dealing with simple consumer problems).

2.1 Primary RCT: Job Board

Origins of the RCT and firm recruitment. Hiring is a natural part of growth 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 argues 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. Like a social planner, SEP's mandate is to create economic value. Given most startups fail, SEP is much more interested in helping its strongest firms succeed rather than improving outcomes for its weaker firms.

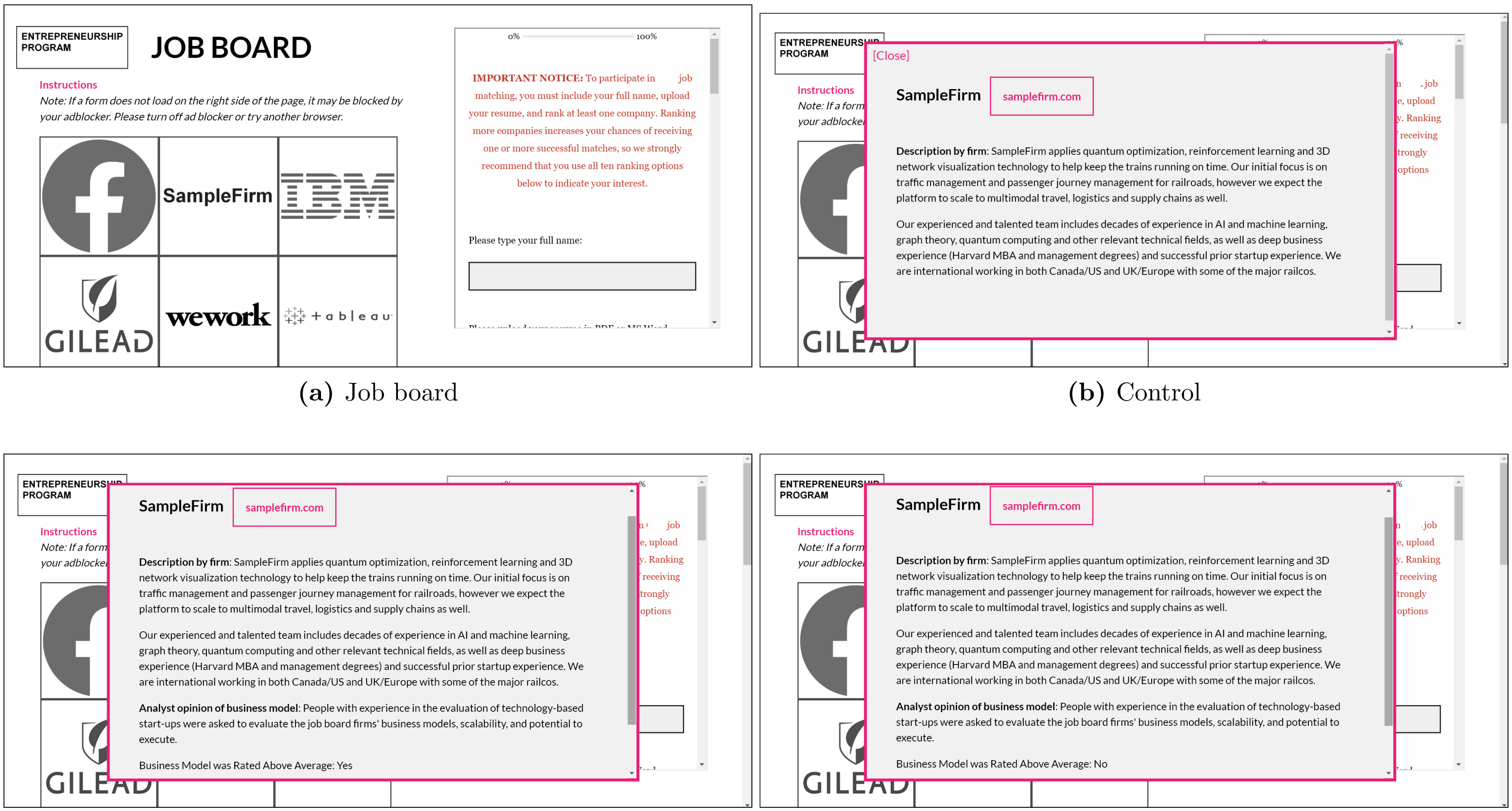
SEP emailed the 183 firms who were part of the 2018-2019 cohort of firms that participated in the program at SEP's headquarter location, and asked if they wished to participate. Of the 183 firms, 26 firms (or 14%) chose to participate. As analyzed later in Section 3, participating firms are broadly representative of SEP firms. Firms were told truthfully that experimentation may occur, but were not informed specifically about the nature of the RCT. Like most startups, SEP firms tend to pay relatively low wages, but offer equity to new hires. To participate, each startup wrote a short ad about their firm.

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 typical SEP founders as two computer science professors with an advancement in image selection for machine learning by autonomous cars, or of a doctor and scientist with a new application of machine learning to healthcare. This is distinct from student startups, where time commitments are limited and who are less likely to grow and hire people.

[7] It wasn't possible logistically to randomize the visual position of firms across workers. This isn't an important concern because our results use *within-firm* differences in whether applicants see expert ratings.

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.

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]

RCT procedure. Potential applicants 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 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: control, information about business model quality, information about science quality, or both. Treatment was assigned by applicant, so each applicant saw expert ratings for all firms or for none. Applicants were unaware that other applicants may see different information or that they would be part of an RCT.

Since workers accessed the job board via a custom link, we can match the randomization each received to their name and ensure that the treatment 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 workers' 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 firm. The stated reason was that, since these firms are small, SEP did not want to overburden founders with hundreds of resumes. 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 rounds. In each round, the algorithm goes through workers in order, matching workers with their highest-ranked firm with 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,

[8] From the 1st business school, we contact 5,681 undergraduate alum (graduated from 2008-2019) and 7,894 graduate alum (graduated from 1946-2018). From the 2nd b-school, we contact 3,701 undergrad alum (graduated from 2009-2015) and 2,083 grad alum (from 2009-2019).

we explained that an award-winning algorithm ensured that it was in the best interest of candidates to truthfully rank firms in order of 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 D for screenshots of how this appeared to subjects]

Ten application cap. 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 3 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, indicating that the constraint does not drive our findings. As discussed in Section 3 ("Worker selection into the RCT"), there is no evidence that workers found the ten application cap to be strange or that it limited who wished to participate.

Worker beliefs. After submitting their ranked list of firms they wished to apply to, workers 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 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

[9]Past work shows that RSD often yields incentive-compatible choices (Chen & Sönmez, 2002; Artemov *et al.*, 2017). We expect our high-skill subjects have an easier time understanding it than most. RSD is incentive-compatible for interest in *getting an interview*, not in getting a job, as workers may worry that better firms get better applicants, and thus the chance of getting a job conditional on an interview is lower. We view this as a feature and not a bug of our RCT, as this issue would occur in a full-scale policy rollout where all workers see ratings. RSD is used around the world in many job markets (Fadlon *et al.*, 2024).

[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 understanding is that firms and workers in the RCT saw the 10-firm constraint as very natural.

$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 D 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. Following past applied work using this scoring rule (Hoffman, 2016), we explain that the incentive system makes it optimal to state one's true beliefs, and we provide math formulas separately for people to look at if interested. [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 firm by a scientist with expertise in the core technology (e.g., a machine learning startup is evaluated by a scientist with expertise in machine learning), as well as detailed written materials provided by the firm before the interview. The scientists are paid to do the assessments as part of their salary from the National Research Council, and thus, take the assessments quite seriously.[12]

When onboarding scientists to conduct assessments, SEP tells scientists to 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 science quality. A binary version of the rating was used to make the ratings easy to interpret for jobseekers, and coarse ratings are common for many ratings (e.g., pass/fail for health inspections).[13]

[11] 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 to report one's true confidence. Appendix B.2 further discusses past work supporting the reliability of eliciting beliefs this way.

[12] National Research Council is the main research institute in Canada's federal government. SEP and National Research Council have an agreement for assessments to count toward one's salary.

[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 startups, SEP required that information

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 when firms apply to SEP. However, since the job board occurred a year after the business model was first evaluated, and firms often update their 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 by venture capitalists and are a standard way of measuring a startup's potential (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, we exploit the fact that science ratings have been conducted on SEP firms for several years. Table 1 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 ("pp") 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

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 are analogous to portfolio managers at venture capital firms who narrow the list of possible investments for partners. In fact, SEP staff often go work for venture capital firms as portfolio managers.

Table 1: 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 regressing start-up success outcomes 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 having a PhD founder, and technology fixed effects.

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 promising business model, could be added to the incubator's employment website. Job 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 & implementation. The job board began in May 2019. Emails were sent in 3 batches (batch 1 = MBA alum of 1st business school, batch 2 = undergrad alum of 1st b-school, batch 3 = MBA and undergrad alum of 2nd b-school), as detailed in Appendix E. For each batch, the 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 firm, as measured by the RSD mechanism (Appendix B.3 gives 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 follow startups through August 2022 to track 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 5-10 firms. Time stamp data indicate candidates took the process seriously. As seen in panel (a) of Figure 1, jobseekers land on a webpage where they can browse 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 = 12 mins, 75th percentile = 54 mins).

2.2 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 evaluated 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. For each firm, MBA students filled out a quantitative evaluation and answered 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, 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 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 were not identical.[15]

Instead of providing a ranked ordering of firms using incentive-compatible RSD, students were asked how interested they would be in working in the firm after graduation on a 1-5 scale.[16] We also elicited the same beliefs about firm outcomes and the perceived quality of science and business as in the primary RCT, and using the same procedure (i.e., using incentives for beliefs about firm outcomes). This allows us to analyze worker beliefs using pooled data from both RCTs.

2.3 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 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

[15] In Appendix F, we provide examples of a firm dossier and expert ratings shown to MBA students.

[16] 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.

or business quality) matter. We turn to these next.

3 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 (available at www.socialscienceregistry.org/trials/4242), we specified that we would consistently analyze four functional forms:

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 gets 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 normalize 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 4.2.

Empirical strategy. Our pre-registered regressions are as follows:

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

Here, n denotes workers and f denotes a firm that a worker evaluates. 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 firm business quality. Likewise, GotScienceInfo_n measures whether a subject randomly gets information about science quality. The variable GoodBizFirm_f indicates whether firm f is rated positively or not in terms of business quality, while GoodScienceFirm_f indicates whether a firm is rated positively in science.

The controls $\boldsymbol{X}_{nf}$ include firm fixed effects to control for underlying firm quality, as well as any strata dummies for RCTs when we do a stratified randomization. We cluster standard errors by worker, as our treatments are randomized across workers.

Our pre-analysis plan (PAP) also states that worker beliefs and perceptions of firm quality will be secondary outcomes. These are analyzed in the same way as our results

Table 2: 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	0.13	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	

Main notes: This table compares applicant characteristics across treatment groups for the Primary RCT (Panel A) and Secondary RCT (Panel B). "Science info" and "Business Info" in column headers refer to the subjects that received science and business scores.
Panel A: In the Primary RCT, randomization was stratified on gender, city, and year of graduation at the time potential applicants were contacted. Other variables were not observable prior to application. 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.
Panel B: In the Secondary RCT, randomization was not stratified.

on worker application rankings. Finally, the PAP specifies that we will run heterogeneity analyses based on whether workers have a STEM degree or not.[17]

Randomization check. Table 2 shows that the four treatment groups of the Primary RCT (Panel A) and Secondary RCT (Panel B) are balanced on observables.

In Panel A, characteristics are only from workers who apply 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 10 years of experience on average and nearly half have an undergraduate STEM degree. 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

[17] The PAP also says we will try to analyze heterogeneity based on firm characteristics related to technological sophistication. In practice, most firms are highly sophisticated technologically. It is not obvious how to compare the sophistication of, e.g., an AI fintech company to a quantum drug discovery platform.

still vary. There are no statistically significant differences across treatment arms in gender, applicant location, graduation year, startup work history, STEM background, or years of work experience. The Secondary RCT's randomization was not stratified, but participants are balanced on race and gender, as seen in Panel B.

Worker selection into the RCT. Based on rich tracking data, among the 587 recipients who open the email and click on the job board link, 43% or 250 workers participate (i.e., apply to at least one firm), suggesting that our job board was well-received and intuitive to workers. Our initial sample is the 19,359 alumni emailed from our partner business school lists, for an overall participation rate of 1.3%. Initially, 37% (or 7,083) of the alumni open the invitation email, reflecting possible spam filters or that alumni may not regularly check old emails on file with the advancement office.[18] Of those who open the email, 8% (or 587) of workers click on the job board link. The 8% rate reflects that many of the alumni contacted are highly established in the business world, and are not actively seeking employment or are not interested in working at a startup.

Appendix Table A1 shows a linear probability model where overall 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.[19] For classes graduating in 1980 or before, the participation rate is 0.46%. Participation increases with class year of graduation, and is highest for the class of 2019, where the rate is 4.3%. Men are 0.9pp more likely (i.e., twice as likely) to participate than women. People in the city where SEP is headquartered are more likely to participate. The treatment 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. These predictors of participation are intuitive, e.g., recent alum are less likely to have well-established careers than older alum.

A critical concern about non-response in RCTs is whether treatment affects participation. This is not an issue for us, given that subjects have identical experience until they access the website, and participation is statistically identical across arms.

Another concern is that the treatment effect on RCT participants would differ from that on a non-experimental job board. This concern is massively reduced because workers don't know they are part of an RCT, i.e., it is a natural field experiment (Harrison & List, 2004). Still, might overall participation patterns indicate something unusual about the target

[18]The average email address was seven years old. At one business school, 46% of emails had university domain names, which seem less likely to be regularly checked than personal emails. This share with university domains is unknown for the other business school. Our email lists are high-quality and not deficient. Alumni email lists often have outdated emails, especially for older alumni.

[19]This participation regression is run on the overall sample of 19,359 alumni. Note that $37\% * 8\% * 43\% = 1.3\%$. Due to university privacy rules, we lack personal IDs from the tracking data, so we cannot run participation regressions conditional on opening the email or clicking the job board link.

sample or the RCT job board? MBA alum are a natural population for analyzing hiring at science-based startups, as such startups seek high-skill workers. Most selection occurs based on opening the email and clicking on the link for the job board. Conditional on clicking the link in the email for the job board, participation is high. In email comments we received and anecdotal discussions with potential participants, the consensus was that the job board—including the ten application cap and the RSD mechanism—was very natural. While most job boards don't have a ten application cap, alumni were given a clear reason for the cap (as mentioned above, to avoid overburdening founders with lots of resumes). Applicants found this reason to be sensible.

Firm selection into the RCT. As noted above, 14% of contacted firms agreed to participate. This 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 are generally weak predictors of whether a startup participates in the RCT, suggesting that participating startups are broadly representative of SEP firms.[20]

4 Results

4.1 Impacts on Job applications

Table 3 shows that expert ratings create large shifts in applications toward better-quality firms. For brevity, we focus on the results in column 3 of Table 3, the specification that simultaneously analyzes both business and science rating treatments. We also provide a visual summary of our results in Figure 2.

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

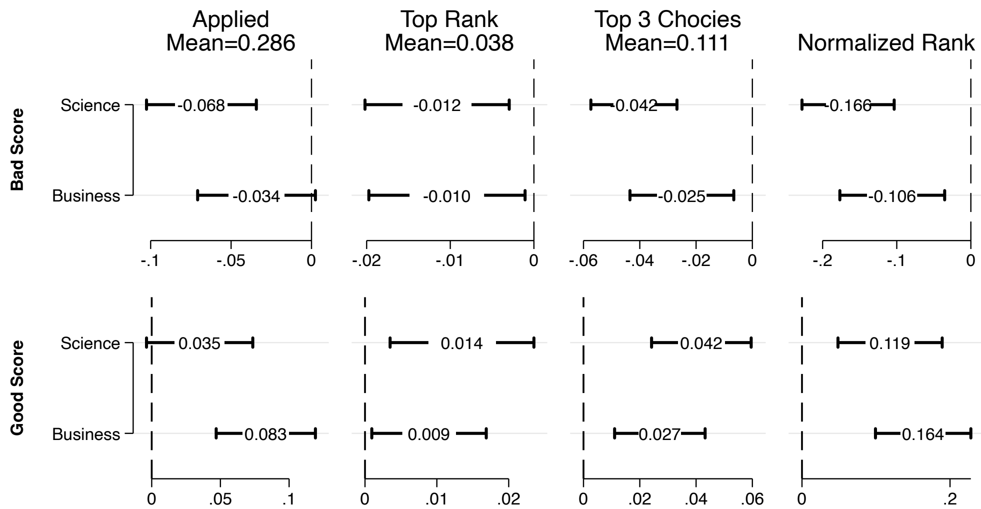
[20]While not available for all non-participating firms, average science quality is similar between participating and non-participating firms. Business quality is unavailable for non-participating firms.

Table 3: The Effect of Expert Ratings on Job Applications

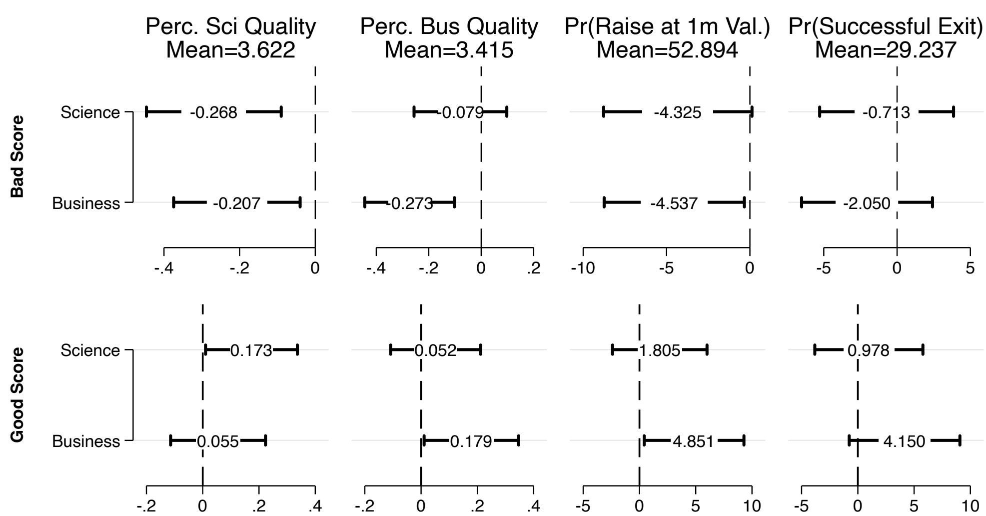
	(1)	(2)	(3)
Panel A: Dep. Var. = Applied			
Science info X Good science	0.102***		0.103***
	(0.025)		(0.025)
Science info	−0.067***		−0.068***
	(0.018)		(0.017)
Business info X Good business		0.116***	0.117***
		(0.026)	(0.026)
Business info		−0.034*	−0.034*
		(0.019)	(0.019)
F(Sci + Sci X GoodSci = 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 in the Primary RCT. Dependent variable for each panel shown at the beginning of the panel. Regressions include startup fixed effects and randomization strata based on gender, whether the worker lives in the same city as SEP headquarters, and 3 bins for year of graduation. Standard errors clustered by worker in parentheses.

Figure 2: The Effect of Expert Ratings on Job Applications and Beliefs: A Visual Summary



(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 2a plots the estimated effect of expert ratings on job applications from the Primary RCT. Figure 2b 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 the 95% level.

Table 4: 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, 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, and so on.

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 32% on a base rate of 3.8%), whereas receiving positive information increases it by roughly an equal amount. We see 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, while positive science information increases the normalized rank by 0.12σ. Negative business information decreases normalized rank by 0.11σ, but positive business information increases normalized rank by 0.16σ.

In sum, applications respond strongly to expert ratings, both business and science ones. Effects tend to be roughly symmetric for negative and positive information.

While Table 3 analyzes the decision to apply to particular firms, exploiting randomized within-firm variation, it is also illustrative to look at the distribution of firms applied to across treatment arms. This is done in Table 4. 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.

Multiple hypothesis testing. Multiple hypothesis testing is an important issue in experiments (List *et al.*, 2019). To account for multiple hypothesis testing related to our four outcomes, Table A7 shows family-wise error rate adjusted p-values based on Westfall & Young (1993). These findings support our main results in Table 3, and show that our conclusions are not driven by multiple testing.

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 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. Across all four outcomes, we see consistent evidence that they are substitutes. This may occur if workers update favorably about business quality based on positive science ratings, and visa versa, as shown below in Table 5 and Section 4.3. Thus, providing both sets of ratings provides less than double the average impact of providing one set of ratings.

4.2 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.[21] 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

[21] 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.

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.[22]

4.3 Beliefs

Beliefs descriptives. Before showing impacts of information on beliefs, we first 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 in beliefs. There is also 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 chance 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 many experienced workers for whom overestimation is perhaps more surprising than in less sophisticated workers.

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.[23] 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 chance of startup success. A likely explanation is that prior experience calibrates the expectations of startup exit.[24]

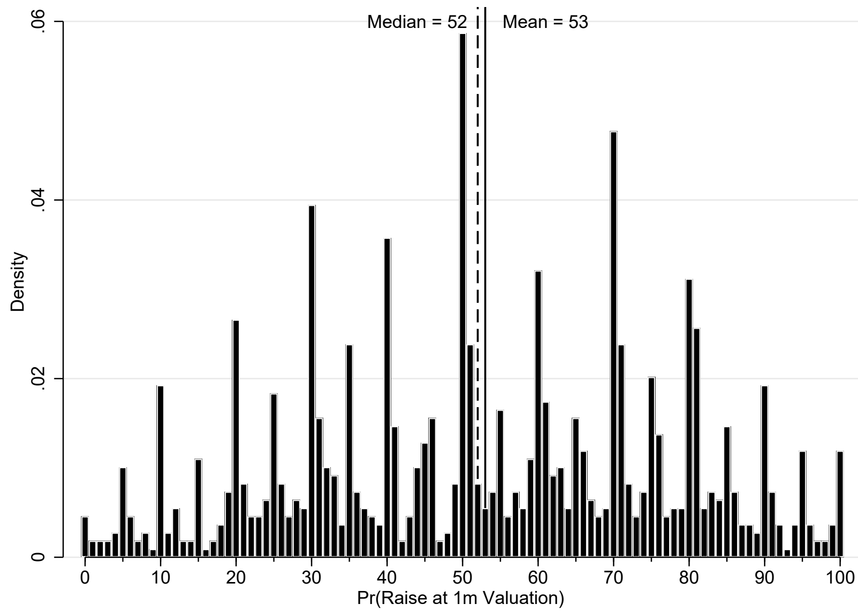
Overconfidence about positive startup outcomes occurs widely across many types of workers and appears highly robust. One concern is that results could be driven solely by the beliefs of unsophisticated applicants. However, as noted above, overoptimism is highly prevalent among workers rated as high-quality and among workers who have STEM degrees.

[22] For example, as of 2019, Google accepted about 0.2% of candidates. `https://cnb.cx/3H29TbB`.

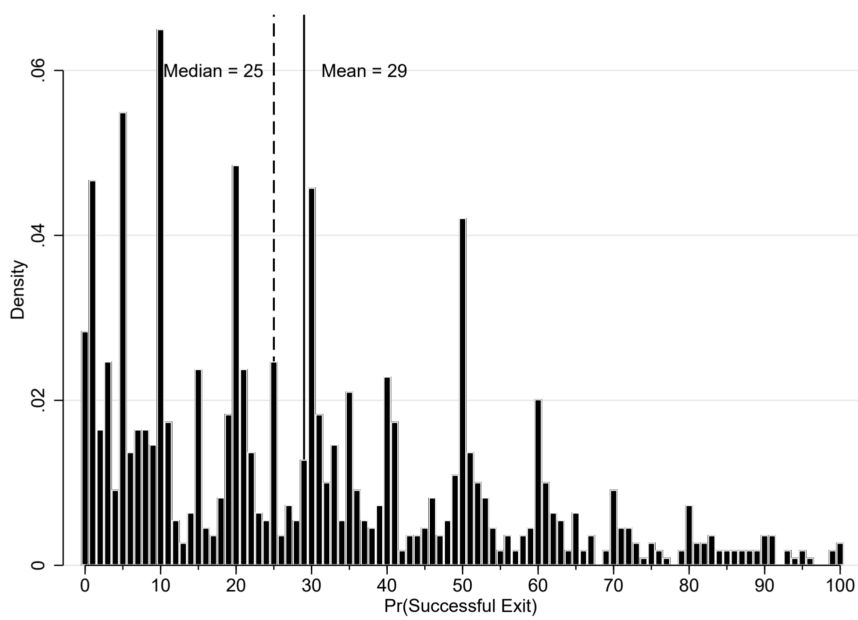
[23] While research often finds that men are more overconfident than women, there are also many situations where men and women appear equally overconfident (e.g., Huffman *et al.*, 2022).

[24] 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.

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.

A second concern is that since 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. However, the finding of overprediction is highly robust to excluding beliefs in multiples of 5 above zero, indicating that overprediction is not driven by rounding errors. Finally, beliefs about firm success are correlated with application decisions (Appendix Table A13), suggesting that beliefs questions are answered seriously.

Impacts of expert ratings on beliefs. Table 5 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 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 2 shows these results graphically, plotting impacts of business and science information on perceived firm quality and chances of positive longer-term outcomes.

4.4 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. We find 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

Table 5: The Effect 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. Dependent variable for each panel shown at the beginning of the panel. Standard errors clustered by worker in parentheses.

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:

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

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.

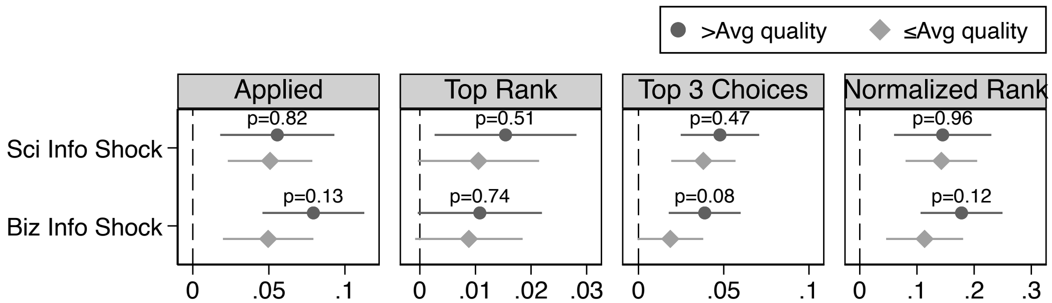
Measuring worker quality. Worker quality is a key variable in models of sorting (Eeckhout & Kircher, 2011), but is not easily observed. To measure it, we contracted with an independent human resources consultant who specializes in startup hiring, having over 15 years of experience, and asked her to rate 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) of Figure 4 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 providing expert ratings could cause adverse selection, and hence misallocation of human capital. Appendix Table A9 shows the same estimates in tabular form.

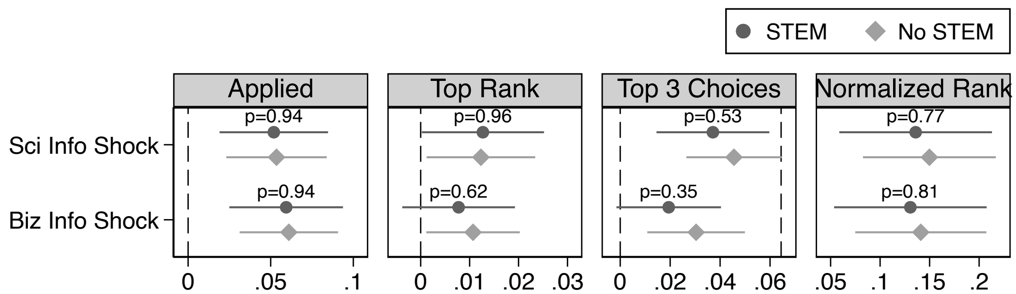
Our estimates are precise enough to usually rule out that higher quality workers respond less than lower quality workers by even relatively small margins in response to business ratings. For the key outcome of applying, with 95% confidence, we rule out that the treatment effect for higher quality workers is more than 18% below that for lower quality workers.[25] Likewise, for the outcomes of top 3 choice and normalized rank, we rule out treatment effect differences favoring lower quality workers of more than 15% and 14%, respectively.

[25] Lower quality workers have a treatment effect coefficient of 0.05. We rule out that higher quality workers

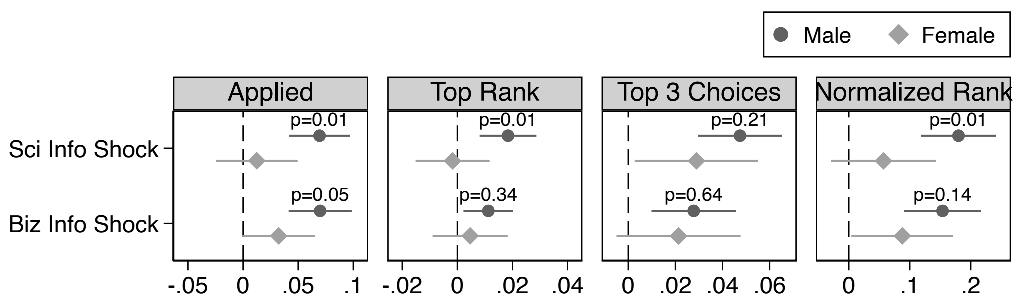
Figure 4: Treatment Effect Heterogeneity by Worker Characteristics



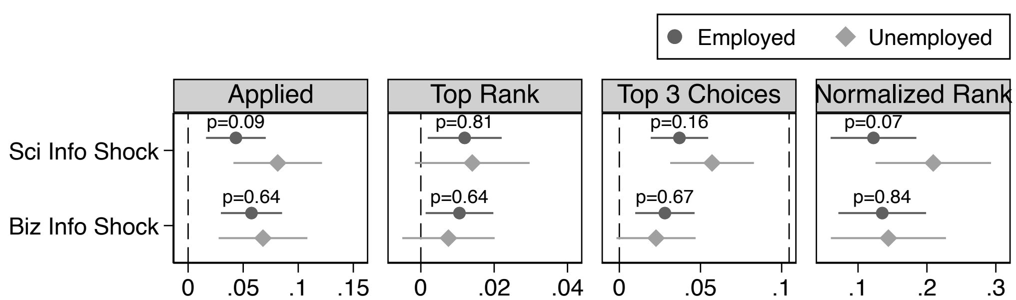
(a) By worker quality



(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 information shocks in the Primary RCT. Estimates are from regressing application outcomes on science and business treatments and their interactions with worker characteristics. Regressions include venture and strata fixed effects. The lines shown are 95% confidence intervals. The p-values correspond to the null hypothesis that there is no heterogeneity in treatment effects based on different worker characteristics.

Panels (b) and (d) of Figure 4 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) of Figure 4 shows that men respond more to science ratings than women, which is robust to adjusting p-values for testing multiple dimensions of worker heterogeneity (an issue stressed in List *et al.* (2019)), as seen in Table A10. There are various explanations for this consistent with prior work on job search, including that women may be more sensitive to commuting and relocation costs (Le Barbanchon *et al.*, 2021). For example, a woman might be interested in applying to a highly-rated firm, but be constrained to do so because the job is far away, whereas a man may be less constrained in relocating (Benson, 2014).[26]

As seen in Appendix Table A11, machine-learning based sorted effects models yield the same qualitative conclusions as those in Figure 4.

Firm characteristics. Figure A3 shows limited evidence of heterogeneity with respect to firm characteristics. There is some weak 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, Table A11 Panel B shows similar findings to Figure A3.

Panel (a) of Figure A3 examines heterogeneity based on whether founders have prior business development experience, such as serving as an executive at another firm. 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 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 business quality). 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.

have a treatment effect coefficient of less than 0.04, based on the confidence interval on the interaction term.

[26]More broadly, it could be that women value non-pecuniary aspects of jobs more than men, and that expert ratings 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). 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 women 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 this would affect our ability to detect significant interaction terms. As seen in panel (c) of Figure 4, all eight point estimates of effects for men are larger than those for women.

Panels (c) and (d) examine heterogeneity based on whether a firm has positive revenue (many science-based startups take time to make revenue) or has external financing. We do not find consistent evidence that effects depend on these factors.

5 Discussion

5.1 Threats to Validity

Hawthorne Effects. A common concern in RCTs is whether results are driven by experimental demand effects, also called Hawthorne Effects. Hawthorne Effects are believed to be most likely when there is an experimental manipulation that subjects are aware of, and where subjects wish to please or influence the researcher (Levitt & List, 2011). However, participants in our study did not know that there was experimental manipulation of what they observe. As noted earlier, workers did not know they were in an RCT, and we received no indication that expert ratings seemed artificial.[27] 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 provide 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

[27]Likewise, 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. Moreover, the RCT treats workers and there are no firm decisions to analyze, so Hawthorne Effects with respect to firms are not an important concern.

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? As discussed in footnote 9 and B.2, past research indicates that these mechanisms are generally effective and reliable. 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. Measurement error in the quadratic scoring rule would not affect our main results; it would also not bias our results on beliefs in Table 5, and will instead contribute to larger standard errors.

External validity. Our study focuses on high-skilled workers applying to science-based startups. Science-based startups are a growing sector, especially given the rise of artificial intelligence, and high-skilled workers are natural to study for this; thus, we believe that our sample is well-suited for our research question, policy-relevant, and independently interesting. That said, it is not clear whether our results on substantial information frictions would 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 (Figure 4), as well as across firm characteristics (Figure A3), suggesting that our effects may also hold among broader populations of workers and firms. While we acknowledge that our results are specific to high-skill workers, we speculate our results may provide a lower bound for a broader population of workers if lower-skill workers are less sophisticated in their ability to evaluate firms.

5.2 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, i.e., 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% are full professors, 21% are associate, and 26% are assistant professors. The remainder are PhD students/postdocs (15%) and industry economists (8%). We also distributed the survey on the Social Science Prediction Platform, where it received 14 additional 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 G provides exact survey questions and other details.

Table 6 summarizes the results of the economist survey. Column 2 of Table 6 lists actual RCT results, whereas Column 3 summarizes economist predictions.

Underprediction. As seen in rows 1-4 of Table 6, economists severely underestimate the quantitative magnitude of our effects. In the absence of expert ratings, economists are relatively accurate in forecasting proportional number of applications received by good firms (i.e., firms rated above-average in terms of both science and business) compared to bad firms (rated below-average in terms of both science and business), as seen in row 1. However, as seen in row 4, when both ratings are shown, good firms get 80% more applications than bad firms, while the median expert prediction is 25% and the mean is only 36%. 86% of economists underpredict the magnitude of the treatment effect, and 60% of economists underpredict the treatment effect by at least half. Substantial underprediction is also observed when economists are asked about providing only science ratings or business ratings.

Other aspects of effects. Economists also fail to predict many aspects of the effects. Our RCT finds comparable effect sizes between science and business ratings, as well as between positive and negative information.[28] However, for each finding, fewer than 15% of economists predict correctly, while most economists (59% and 72%, respectively) incorrectly believe that business rating and negative information would yield a larger impact.

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, for each of these findings, less than 30% of economists predict correctly. The most popular economist response is to predict stronger responses for STEM degree and high-quality workers. While we find larger treatment effects for men than women, only 14% of economists predict this.

[28]Rows 2-3 of Table 6 show that science and business ratings increase 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 5 of Table 6 shows that science or business ratings had similar effects (based on Table 3)? This reflects that the row 5-7 findings are based on overall effects using all the data, whereas rows 1-4 focus solely on firms that have both good science and good business or that have bad science and bad business.

Table 6: Economist Expert Survey Results: Economists Underestimate the Impact of Ratings and Fail to Predict Other Aspects of our Treatment Effects

Prediction problem	Actual	Expert Prediction
What percent more/less apps to good sci & biz firms compared to bad sci & biz firms when:		
1. No ratings shown	11%	Mean = 15 p25=3.5%, p50=10%, p75=20%
2. Science ratings shown	32%	Mean = 23 p25=10%, p50=19%, p75=30%
3. Business rating shown	73%	Mean = 28 p25=10%, p50=20%, p75=40%
4. Both ratings shown	80%	Mean = 36 p25=12%, p50=25%, p75=50%
Was the effect larger for:		
5. Science or business rating	Similar effect	Larger for sci 22%, Larger for biz 59% **Similar Effect 15%**, No effect 4%
6. Positive or negative rating	Similar effect	Larger for +ve 14%, Larger for -ve 72% **Similar effect 11%**, No effect 3%
Were science/business ratings:		
7. Complements or substitutes	Substitutes	Complements 67%, **Substitutes 27%**, No effect 6%
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, 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. The exact survey questions and information shown to responders are provided in Appendix G. In terms of order in the survey, economists first made predictions related to 5-7, followed by 1-4, followed by 8-10.

Perhaps many economists fail to predict various aspects of our effects because some of our heterogeneity estimates are somewhat imprecise? This seems very unlikely to be a complete explanation. As also observed in Englmaier *et al.* (2024), expert beliefs are quite dispersed. Consider heterogeneity by whether jobseekers have a STEM degree. Expert predictions are roughly evenly split over three options (smaller for workers with a STEM degree, no difference, larger for workers with a STEM degree). If roughly 1/3 of experts each predict one of three options, then 2/3 of experts will be incorrect in their predictions.

5.3 Why Don't Firms Provide More Quality Signals?

If asymmetric information about firm quality is severe, why don't high-quality firms give credible quality signals? Empirically, they do not. In self-written ads, 19 of 26 SEP firms describe technical details, 23 describe their commercial product, and 8 mention their current or planned business model. However, only 4 of 26 give *any* credible quality signal.[29] That so few startups in our RCT provide credible quality signals in their job ads broadly supports that the findings of the RCT were not *ex ante* obvious to startups (in addition to not being obvious to economists in Section 5.2).

Our primary explanation is founders do not realize that without credible signals, their firm is hard for applicants to evaluate. Alternatively, even top startups may not have outside credible signals to cite. This isn't the case for the 26 firms in our primary RCT: we verified in SEP internal documents that all 26 had outside signals as defined above which could've been shown in a job ad.

A third explanation is our sample is unusual. It is, after all, a small sample of startups that are science-heavy. To investigate this pattern more broadly, we examine the content of ads 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).[30] From an ad's full text, 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 E.4).

Appendix Table A12 summarizes this AngelList data. Only 23% of ads contain even

[29] We include any mention of founder education, whether the firm is a spinout, whether the firm 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 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.

[30] These are all ads posted between 10/30/2020 and 11/13/2020.

one such signal, though 93% describe the product being sold and 25% 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. Still, it is striking that more than 3/4 of startups do not try to differentiate their firm on quality. Even restricting to non-technical, business development ads (40% of the sample), only 24% include credible outside signals. In sum, the lack of signals which permit workers to sort to high-quality startups appears to be fairly general, not a quirk of SEP.

Why startups may not be optimizing? How is it possible that founders may underpredict the value of providing credible signals? This may seem surprising if firms usually optimize. However, founders may suffer from the same misperception as economist experts, who substantially underestimate the quantitative magnitude of our effects in the expert survey (Section 5.2). Relative to economists, misperception about the value of credible signals in hiring may naturally be larger for startup founders, who often have technical / STEM backgrounds, and who may lack detailed knowledge about labor markets and hiring.

Do firms not provide signals because getting high-quality applicants doesn't matter much? Another possibility is that firms may not provide signals because the value of getting high-quality applicants is low relative to the cost of producing signals. While precise quantification is difficult, our discussions with the startups indicate this to be highly unlikely. The startups in the RCT say they are keen to receive high-quality applicants, consistent with the large qualitative literature in management on the importance and challenge of hiring for startups (Wasserman, 2013), and providing quality signals does not seem onerous.

5.4 Translating Treatment Effects on Quality of Firm Applied To into Effects on Worker Earnings

We estimate here how much a worker's expected lifetime earnings would change if they work for the average *post-treatment* firm they apply to instead of the average *pre-treatment* firm. We estimate that the expected benefit is $800-$2,800 USD per applicant.

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. In math, the benefit to a treated worker is $\mathbb{E}[\Delta_q] \times \frac{\partial R}{\partial q} \times \mathbb{E}[LS]$. Here, $\mathbb{E}[\Delta_q]$ 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 accruing to an early hire.

Both science and business rating treatments shift roughly 9.5% of applications from below- to above-median firms.[31] 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 $1.35 \times .095 = .128\sigma$. A $.128\sigma$ increase in science quality, as shown in Table 1, predicts a 1.2pp increase in the chance a firm raises a venture round after SEP. A similar calculation using just the (noisier) outcomes of firms in the RCT as discussed in Section 2.1 suggests that a $.128\sigma$ increase in science or business quality raises the chance the firm raises money or hires 10+ employees within 3 years of the RCT by .8pp to 2.8pp. 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 pay from working at a better firm, Kerr *et al.* (2013) estimate being funded increases the chance of a positive successful exit by roughly 10pp. Thus, each of our treatments leads the average application to be made to a firm with a .08pp to .28pp 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 rough calculation suggests the equity share of an early employee has an expected value of just under \$1m.[32] Thus, each treatment has an expected value per applicant of \$800-\$2,800 USD, solely in terms of working at a firm that is more likely to have an IPO or be acquired.[33]

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

6 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

[31] Science ratings raise the chance of an above-median application by 12% and reduce the chance 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.

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

[33] 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 how treatments affect choice quality, but differs from the treatments' average earnings benefit (e.g., since treatments shift applications to better firms, workers may also face more competition). Besides more 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.

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. While it may seem surprising that startups don't provide more quality signals on their own, economist experts significantly underpredict the magnitude of our treatment effects.

Policy implications. 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 ratings may be useful and important in addressing workers' informational deficits.

Organizational implications and rollout. Turning to implications of the study for SEP itself, the organization seemed 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 a quick and well-resourced 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.

In response to our RCT results documenting the difficulty high-quality startups have in hiring early employees, in late 2021, SEP rolled out a non-experimental pilot job board for graduating SEP firms looking to hire business school alumni. Interestingly, expert ratings were not initially used on the job board, presumably reflecting that SEP, which is a non-profit, faced financial, staffing, and attention constraints during the pandemic. The pilot job board without expert ratings operated only for one year,[34] but an SEP executive indicated that they are likely to consider it again in the future, and expert ratings are a possibility for the future. In addition to the non-experimental rollout, as of Spring 2023, SEP started providing coaching regarding the importance of quality signals in startup hiring. Startups are encouraged to provide quality signals in job ads when available, even if it feels like bragging. That aspects of the RCT have scaled in a world-leading entrepreneurship program support the external validity of the findings (List, 2020).[35]

[34] It is possible that the lack of expert ratings limited the success of the non-experimental pilot job board that was rolled out. Limited staffing due to Covid may also play a role, according to SEP leadership.

[35] Beyond scaling, List (2020) also discusses the importance of considering selection, attrition, and natu-

Open questions. 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. Could it be that such credible signals do not exist, or are too costly for startups to generate? In our sample, this seems unlikely, as all 26 startups had credible signals they could have provided. 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.

While we examine the imperfect information of workers applying to high growth startups, and the extent to which even coarse information about firm quality changes their job search behavior, it is an open question how widespread this inefficiency is in the broader labor market. There are many aspects of a job workers imperfectly observe, even at large firms: how quick are promotions, do bosses train young workers, does the firm have a bright future, and so on. Workers can of course rely on first-hand testimony from friends or commenters on sites like GlassDoor, or infer prospects from the salaries they are offered. Nonetheless, just as how a large literature now shows inefficient job matching due to firm uncertainty about worker quality, the reverse (i.e., worker uncertainty about firm quality) also seems important, and is currently not well understood.

References

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

Aran, Yifat, & Murciano-Goroff, Raviv. 2023. Equity Illusions. *Journal of Law, Economics, and Organization*, Forthcoming.

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.

Ashraf, Nava, Bandiera, Oriana, Minni, Virginia, & Quintas-Martinez, Victor. 2022. Gender Roles and the Misallocation of Labour Across Countries. *Mimeo, LSE*.

Bapna, Sofia, Benson, Alan, & Funk, Russell. 2021. Rejection Communication and Women's Job-search Persistence. *Available at SSRN 3953695*.

ralness. We discuss selection into the RCT by workers and firms in Section 3. There is no subject attrition in our RCT, and the RCT is highly natural since workers don't know they are in an experiment.

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. *AEJ Macro*, **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, & Shaw, Kathryn. 2024. *Thinking About What Managers Do.* Working Paper, Stanford University.

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.

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. *QJE*, **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: Handbook of Economic Expectations*. Academic Press.

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

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. *ECMA*, **86**(6).

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

Dustmann, Christian, Puhani, Patrick A, & Schönberg, Uta. 2017. The Long-term Effects of Early Track Choice. *The Economic Journal*, **127**(603), 1348–1380.

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.

Englmaier, Florian, Grimm, Stefan, Grothe, Dominik, Schindler, David, & Schudy, Simeon. 2024. The Effect of Incentives in Nonroutine Analytical Team Tasks. *Journal of Political Economy*, **132**(8), Forthcoming.

Fadlon, Itzik, Lyngse, Frederik Plesner, & Nielsen, Torben Heien. 2024. *Early Career Setbacks and Women's Career-Family Trade-Off*. 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? *J. Financial Econ.*, **135**(1), 169–190.

Harrison, Glenn W., & List, John A. 2004. Field Experiments. *Journal of Economic Literature*, **42**(4), 1009–1055.

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.

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

Huffman, David, Raymond, Collin, & Shvets, Julia. 2022. Persistent Overconfidence and Biased Memory: Evidence from Managers. *AER*, **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. 2024. Worker Beliefs About Outside Options. *Quarterly Journal of Economics*, Forthcoming.

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

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

Leibbrandt, Andreas, & List, John A. 2018. *Do Equal Employment Opportunity Statements Backfire? Evidence from a Natural Field Experiment on Job-Entry Decisions*. Tech. rept. National Bureau of Economic Research.

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

LEVITT, STEVEN D., & LIST, JOHN A. 2011. Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, **3**(1), 224–238.

LI, XIAOMIN, & CAMERER, COLIN F. 2022. Predictable Effects of Visual Salience in Experimental Decisions and Games. *QJE*, **137**(3), 1849–1900.

LIST, JOHN A. 2020. *Non est disputandum de generalizability? A glimpse into the external validity trial.* Tech. rept. National Bureau of Economic Research.

LIST, JOHN A., SHAIKH, AZEEM M, & XU, YANG. 2019. Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics*, **22**, 773–793.

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.*

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.

STAIGER, DOUGLAS O., & ROCKOFF, JONAH E. 2010. Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives*, **24**(3), 97–118.

STERN, SCOTT. 2004. Do Scientists Pay to be Scientists? *Mgmt. Science*, **50**(6), 835–853.

WASSERMAN, NOAM. 2013. *The Founder's Dilemmas.* Princeton University Press.

WESTFALL, PETER H, & YOUNG, S STANLEY. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment.* Vol. 279. John Wiley & Sons.

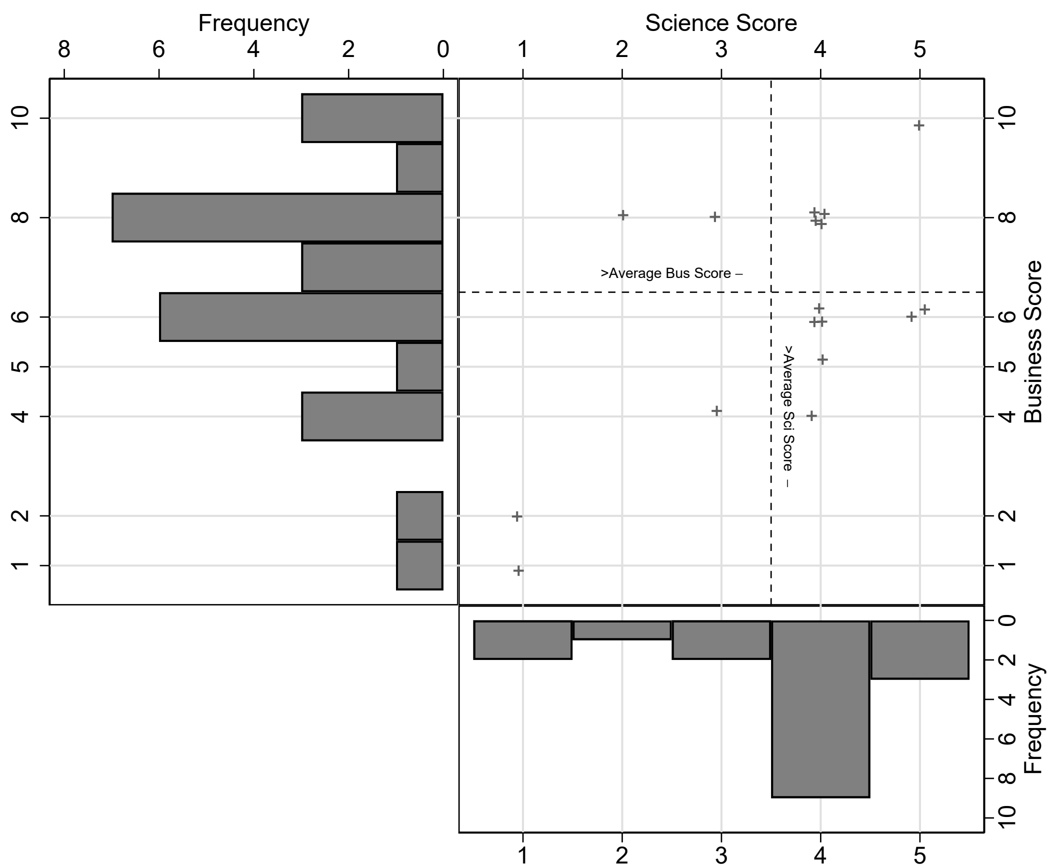
WISWALL, MATTHEW, & ZAFAR, BASIT. 2015. Determinants of College Major Choice: Identification using an Information Experiment. *Rev. Econ. Studies*, **82**(2), 791–824.

Online Appendix: "Information Frictions and Employee Sorting Between Startups"

The Online Appendix consists of the following parts. Appendix A provides additional figures and tables. Appendix B provides further discussion on different parts of the main text. Appendix C provides the Theory Appendix. Appendix D provides screenshots from the Primary RCT. Appendix E is a Data Appendix containing definitions of variables and details on the creation of the sample. Appendix F provides key documents from the Secondary RCT. Appendix G provides further details on the survey of economist experts. Appendix H shows the detailed explanation of the quadratic scoring rule that was made available to subjects (in addition to the simpler and intuitive explanation that was provided to subjects and that can be seen in Appendix D).

Appendix A Additional Figures and Tables

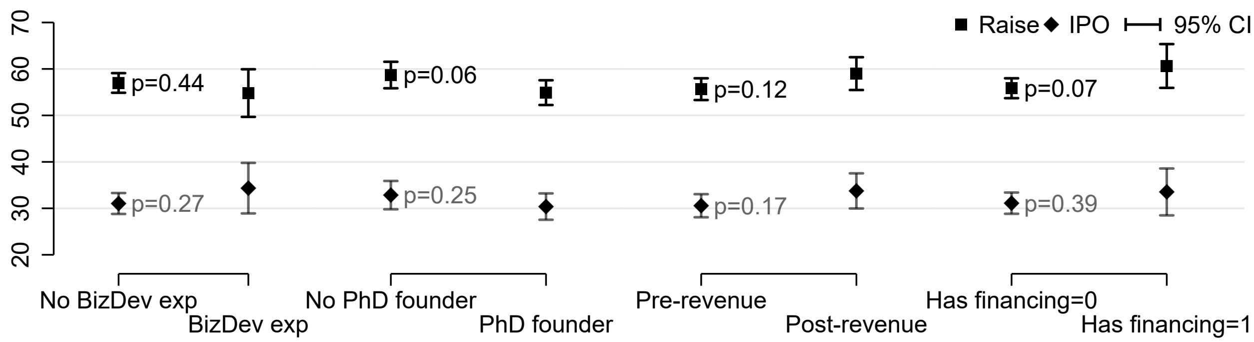
Figure A1: Distribution Plots of Science and Business Scores



Notes: This figure shows the distribution of startup scores in the Primary RCT, excluding 9 startups that were missing the science score. Scatterplot points are jittered for clarity.

Figure A2: Distribution of Worker Beliefs for Raise and Exit in 1 Year by Worker and Firm Characteristics

(a) By worker characteristics



(b) By firm characteristics

Notes: This figure shows the mean and 95% confidence interval of the Primary RCT incentivized beliefs about the probability of successful funding (square symbols) and successful exit (diamond symbols). Difference in means test p-values are reported beside symbols. For raise, workers are asked "What is the probability that the firms below raise money at a valuation of at least CAD\$1,000,000 within 1 year of the time this information was prepared?" For exit, workers are asked "What is the probability that the firm in question has an initial public offering (IPO) or is acquired at CAD\$50,000,000 or more within 1 year of the time the information was prepared?"

Table A1: Selection of Alumni into the Primary RCT

	Submitted job app
Male	1.030***
	(0.163)
City is SEP HQ	0.758***
	(0.183)
Graduation Year, Base Level = 1980	
1985	0.402
	(0.455)
1995	0.392
	(0.409)
2005	1.110***
	(0.404)
2013	1.152***
	(0.352)
2018	0.867**
	(0.360)
2019	4.560***
	(0.771)
Treatment Group, Base Level = No Info	
Business + Science info	−0.037
	(0.233)
Business info	0.025
	(0.235)
Science info	−0.268
	(0.223)
R^2	0.01
Observations	19,359

Notes: This table examines overall selection into the Primary RCT. It shows a linear probability model, where RCT participation (defined as applying to at least one firm on the job board) is regressed on subject characteristics. Coefficients are multiplied by 100 for readability. Robust standard errors in parentheses. An observation is an alumni who is emailed. More details on the selection process are provided in Section 3 of the main text.

Table A2: Selection of Firms into the Primary RCT

	Firm Characteristic Means by RCT Treatment Arm					
	Bad Biz Bad Sci	Bad Biz Good Sci	Good Biz Bad Sci	Good Biz Good Sci	Non-RCT Firm Means	All Firms, Selection Regression
Has financing	0.0	0.1	0.2	0.4	0.2	-0.028
						(0.061)
Num. employees	0	7	0	3	4	-0.002
						(0.006)
Num. founders	3	3	2	3	2	0.044*
						(0.025)
PhD Founder	0.4	0.7	0.3	0.6	0.5	-0.013
						(0.054)
BizDev exp	0.2	0.0	0.6	0.2	0.5	-0.109*
						(0.055)
Female founder	0.6	0.3	0.2	0.2	0.2	0.017
						(0.063)
Log(Revenue)	0.00	3.92	3.23	6.89	3.57	-0.000
						(0.005)
Log(Capital)	11.35	6.94	4.72	9.52	6.88	0.003
						(0.004)
R^2						0.05
Observations	5	7	9	5	157	183

Notes: This table shows descriptive statistics for the 26 startups in the Primary RCT (Job Board), and the remaining 157 startups in the same SEP cohort; these 183 firms make up the full cohort of 2018-2019 firms who participated in streams at SEP's primary location. The first four columns present means of variables in the four RCT treatment arms in the Primary RCT. The fifth column presents means for the firms who chose not to participate in the RCT. The final column presents results from a selection regression, where the dependent variable is whether a startup chose to participate in the Job Board (0 or 1), and with robust standard errors in paretnehses. As can be seen, observable characteristics are generally weak predictors of whether a startup participates in the job board.

Table A3: Share of Workers by Number of Applications and Treatment Group

	Science & Business Info	Business Info	Science Info	No Info
Number of Applications Submitted:				
=10	0.45	0.52	0.32	0.47
=1	0.02	0.00	0.08	0.02
≤ 3	0.16	0.06	0.19	0.14
≥ 5	0.75	0.91	0.74	0.79

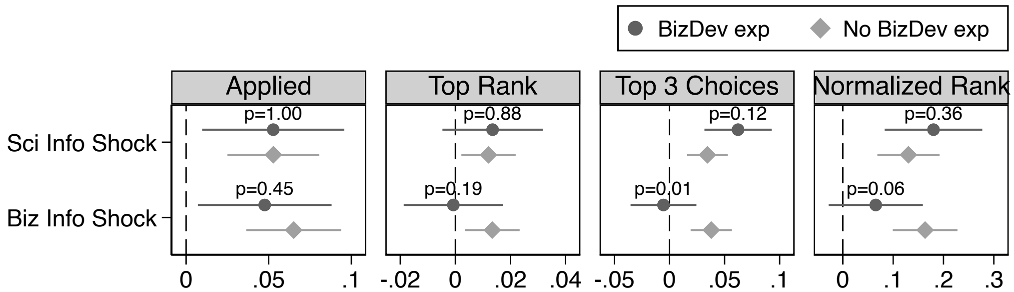
Notes: This table shows the intensity of job applications by different treatment groups in the Primary RCT. Rows show the share of workers in each treatment group who used all, one, less than four, and at least half of the possible application slots by ranking startups among their top ten places to work.

Table A4: Non-Experimental Predictors of Job Applications

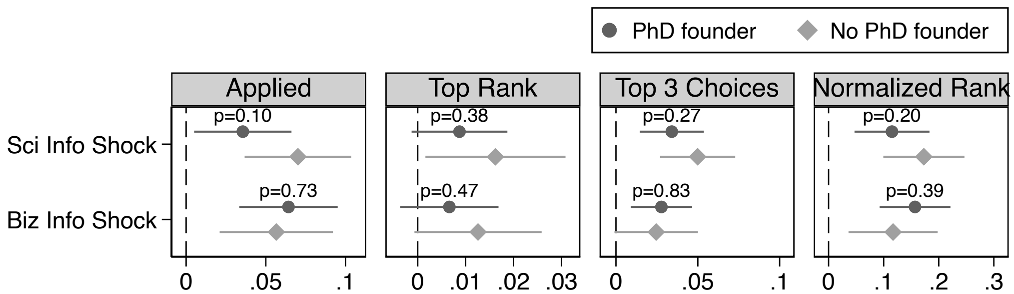
	Applied	Top Ranked	Top 3 Choices	Normalized Rank
Panel A: No Information				
Has financing	−0.071	−0.091**	−0.099*	−0.221
	(0.073)	(0.039)	(0.051)	(0.140)
Num. founders	0.033	0.026**	0.023	0.060
	(0.021)	(0.012)	(0.015)	(0.041)
Num. employees	−0.002	−0.000	−0.002	−0.008*
	(0.002)	(0.001)	(0.002)	(0.004)
Pct SEP activities completed	0.157*	0.000	−0.012	0.159
	(0.087)	(0.030)	(0.045)	(0.158)
PhD Founder	−0.097**	−0.008	−0.049**	−0.188**
	(0.037)	(0.010)	(0.022)	(0.079)
BizDev exp	−0.008	0.012	0.032	0.028
	(0.034)	(0.016)	(0.023)	(0.068)
Female founder	0.009	−0.059***	−0.023	−0.011
	(0.035)	(0.022)	(0.033)	(0.089)
Log(Revenue)	0.008**	0.005**	0.009***	0.026***
	(0.003)	(0.002)	(0.003)	(0.007)
Log(Capital)	0.002	0.001	0.001	0.005
	(0.002)	(0.001)	(0.002)	(0.005)
Top 1/3 Page	0.036	0.042***	0.031	0.117*
	(0.032)	(0.016)	(0.024)	(0.069)
R^2	0.08	0.08	0.11	0.12
Observations	1,716	1,716	1,716	1,716
Panel B: Full Sample				
Has financing	0.014	−0.059***	−0.045	−0.018
	(0.043)	(0.022)	(0.028)	(0.086)
Num. founders	0.010	0.013**	0.014	0.024
	(0.014)	(0.007)	(0.010)	(0.030)
Num. employees	−0.002**	−0.002***	−0.003***	−0.009***
	(0.001)	(0.001)	(0.001)	(0.002)
Pct SEP activities completed	0.098**	−0.004	−0.005	0.074
	(0.044)	(0.017)	(0.027)	(0.088)
PhD Founder	−0.062***	−0.019***	−0.037***	−0.139***
	(0.017)	(0.006)	(0.011)	(0.037)
BizDev exp	−0.048***	0.001	−0.014	−0.082**
	(0.017)	(0.008)	(0.012)	(0.039)
Female founder	−0.007	−0.034***	−0.034**	−0.040
	(0.020)	(0.010)	(0.015)	(0.044)
Log(Revenue)	0.010***	0.006***	0.009***	0.028***
	(0.002)	(0.001)	(0.001)	(0.004)
Log(Capital)	−0.000	0.000	0.000	−0.001
	(0.001)	(0.001)	(0.001)	(0.003)
Top 1/3 Page	0.022	0.028***	0.036***	0.077**
	(0.017)	(0.008)	(0.012)	(0.037)
R^2	0.05	0.05	0.06	0.07
Observations	6,500	6,500	6,500	6,500

Notes: This table shows non-experimental predictors of job applications. All models include fixed effects for the specialized technology stream of the SEP program to which startups were admitted. Streams are based on core technology or industry, and include machine learning, quantum machine learning, blockchain, space, cities, and health. Standard errors clustered by worker in parentheses.

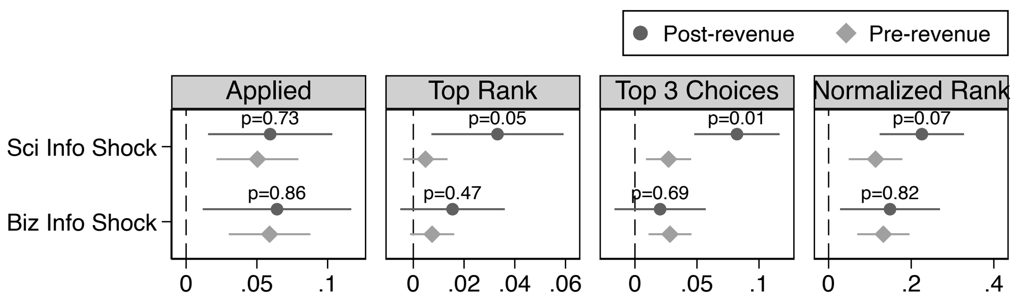
Figure A3: Treatment Effect Heterogeneity by Firm Characteristics



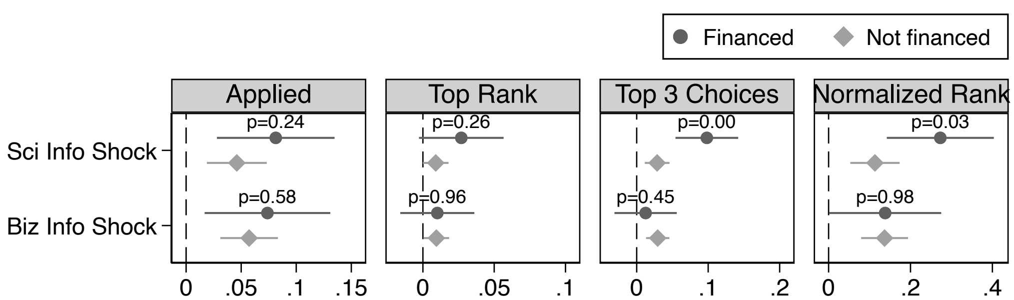
(a) By whether founders have prior business development experience



(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 information shocks in the Primary RCT. Estimates are from regressing application outcomes on science and business treatments and their interactions with worker characteristics. Regressions include venture and strata fixed effects. The lines shown are 95% confidence intervals.

Table A5: Impact of Expert Ratings on Unincentivized Job Interest

	(1)	(2)	(3)
Science info X Good science	0.117		0.118
	(0.125)		(0.125)
Science info	-0.171^{*}		-0.172^{*}
	(0.101)		(0.100)
Business info X Good business		0.235^{**}	0.239^{**}
		(0.118)	(0.118)
Business info		-0.176^{**}	-0.176^{*}
		(0.090)	(0.090)
F(Sci + Sci X GoodSci = 0)	0.535		0.536
F(Bus + Bus X GoodBus = 0)		0.520	0.491
Observations	1,104	1,104	1,104

Notes: This table shows the within-startup effect of information on the candidate's normalized interest in working for the start-up using pooled data from the Primary and Secondary RCTs. Worker interest is a score from 1 to 5 (highest). Standard errors clustered by worker in parentheses.

Table A6: Correlations between Heterogeneity Dimensions and Worker Beliefs

	Raise at $1m Valuation	IPO or $50m Acquisition
Panel A: Worker Characteristics		
>Avg quality	1.702	2.258
	(2.687)	(3.560)
Male worker	1.420	−6.739
	(2.903)	(4.114)
STEM	4.986*	1.272
	(2.804)	(3.570)
Employed	−3.671	−1.400
	(3.346)	(4.540)
R^2	0.02	0.02
Panel B: Firm Characteristics		
BizDev exp	−2.763	3.186
	(3.127)	(3.184)
PhD Founder	−4.157**	−3.382
	(1.836)	(2.055)
Post-revenue	1.982	2.938
	(2.165)	(2.166)
Has financing	5.139*	1.355
	(2.746)	(2.642)
R^2	0.02	0.01
Observations	534	534
Mean of DV	56.64	31.28

Notes: This table shows worker and firm predictors of beliefs about firm success in the Primary RCT. The dependent variables are shown at the top of each column. Worker were asked to submit their beliefs about three randomly selected firms. Data is 604 belief responses from Primary RCT, of which 7 are missing Pr(Raise at 1m Valuation), 4 are missing Pr(Successful Exit), and 59 are missing both. Standard errors clustered by worker in parentheses.

Table A7: Multiple Hypothesis Testing
Multiplicity of Outcomes

Dependent Variables	Applied	Top Rank	Top 3 Choices	Normalized Rank
Science info X Good science	{0.000}	{0.010}	{0.000}	{0.000}
Science info	{0.001}	{0.010}	{0.000}	{0.000}
Business info X Good business	{0.000}	{0.033}	{0.006}	{0.000}
Business info	{0.082}	{0.068}	{0.026}	{0.016}

Notes: This table displays family-wise error rate (FWER) adjusted p-values to account for analyzing the impact of information on multiple outcome variables shown in Table 3, based on Westfall & Young (1993) free step-down procedure (5,000 replications) and while accounting for clustering by worker in bootstrapping. Each p-value adjusts for testing four hypotheses on whether the treatment equals zero for 4 outcome variables. The specification is $y_{nf} = \alpha_0 + \alpha_1 \text{GotBizInfo}_n + \alpha_2 \text{GotBizInfo}_n \times \text{GoodBizFirm}_f + b_1 \text{GotScienceInfo}_n + b_2 \text{GotScienceInfo}_n \times \text{GoodScienceFirm}_f + \boldsymbol{X}_{nf} + \varepsilon_{nf}$. Table A10 provides a multiple hypothesis testing correction related to multiple dimensions of heterogeneity.

Table A8: Tests of Complementarity between Science and Business Ratings

	Applied	Top Rank	Top 3 Choices	Normalized Rank
Biz info	−0.055	−0.019*	−0.022	−0.124*
	(0.035)	(0.011)	(0.021)	(0.072)
Sci info	−0.129***	−0.027**	−0.058***	−0.298***
	(0.035)	(0.011)	(0.020)	(0.068)
Biz Info X Sci Info	0.107**	0.038**	0.053*	0.245**
	(0.051)	(0.015)	(0.029)	(0.104)
Biz info X Good firm	0.137**	0.030	0.045	0.295***
	(0.054)	(0.019)	(0.033)	(0.110)
Sci info X Good firm	0.061	0.022	0.067**	0.228**
	(0.046)	(0.018)	(0.031)	(0.100)
Biz Info X Sci Info X Good Firm	−0.062	−0.014	−0.015	−0.141
	(0.078)	(0.027)	(0.050)	(0.169)
R^2	0.07	0.03	0.04	0.07
Observations	2500	2500	2500	2500

Notes: This table shows tests of complementarity between science and business rating information on job applications. The sample is restricted to good and bad firms, defined as whether the firm is rated as above-average on both dimension or it is not rated as above-average on both dimension. The specification is identical to that of Table 3, except for the addition of the interaction variables *Biz Info X Sci Info* and *Biz Info X Sci Info X Good Firm*.

Table A9: Treatment Effect Heterogeneity by Worker Characteristics

	Applied	Top Ranked	Top 3 Choices	Normalized Rank
Panel A: Heterogeneity by worker quality				
Sci Info Shock	0.051***	0.011*	0.038***	0.143***
	(0.014)	(0.006)	(0.010)	(0.032)
>Avg quality × Sci Info Shock	0.005	0.005	0.010	0.002
	(0.021)	(0.007)	(0.014)	(0.049)
Bus Info Shock	0.050***	0.009*	0.019*	0.113***
	(0.015)	(0.005)	(0.010)	(0.034)
>Avg quality × Bus Info Shock	0.030	0.002	0.020*	0.065
	(0.020)	(0.006)	(0.012)	(0.041)
Panel B: Heterogeneity by whether the worker has a STEM undergraduate degree				
Sci Info Shock	0.054***	0.012**	0.046***	0.150***
	(0.015)	(0.006)	(0.010)	(0.034)
STEM × Sci Info Shock	−0.002	0.000	−0.008	−0.014
	(0.020)	(0.007)	(0.013)	(0.047)
Bus Info Shock	0.061***	0.011**	0.030***	0.141***
	(0.015)	(0.005)	(0.010)	(0.034)
STEM × Bus Info Shock	−0.002	−0.003	−0.011	−0.011
	(0.020)	(0.006)	(0.012)	(0.043)
Panel C: Heterogeneity by worker gender				
Sci Info Shock	0.012	−0.002	0.029**	0.057
	(0.019)	(0.007)	(0.013)	(0.044)
Male × Sci Info Shock	0.057***	0.020***	0.018	0.123**
	(0.021)	(0.007)	(0.015)	(0.049)
Bus Info Shock	0.032*	0.005	0.021	0.088**
	(0.017)	(0.007)	(0.013)	(0.042)
Male × Bus Info Shock	0.038*	0.007	0.006	0.066
	(0.019)	(0.007)	(0.014)	(0.045)
Panel D: Heterogeneity by a worker's current employment status				
Sci Info Shock	0.081***	0.014*	0.057***	0.209***
	(0.020)	(0.008)	(0.013)	(0.043)
Employed × Sci Info Shock	−0.038*	−0.002	−0.020	−0.087*
	(0.022)	(0.008)	(0.014)	(0.048)
Bus Info Shock	0.068***	0.008	0.023*	0.144***
	(0.020)	(0.006)	(0.012)	(0.042)
Employed × Bus Info Shock	−0.010	0.003	0.005	−0.009
	(0.022)	(0.006)	(0.013)	(0.045)

Notes: This table shows heterogeneity in worker response to information shocks in the Primary RCT. It shows the same estimates from Figure 4, but in tabular form. In addition, Figure 4 presents treatment effects for both values of a characteristics (e.g., for men and for women), whereas this table presents interaction effects. Each panel analyzes interaction terms involving a different worker characteristic. There are four regressions presented per panel, each with a different dependent variable shown in the column headers. Each regression includes the characteristic by itself as a regressor, as noted in the regression equation in the main text, but this coefficient is suppressed for clarity. As described in the main text, worker quality is measured by a startup-focused HR expert.

Table A10: Multiple Hypothesis Testing
Multiplicity of Heterogeneity Dimensions

	Worker is male	Worker is high quality
Science info shock	{0.044}	{0.968}
	[0.028]	[1.000]
Business info shock	{0.181}	{0.314}
	[0.201]	[0.392]

Notes: This table displays family-wise error rate (FWER) adjusted p-values in curly brackets (Bonferroni adjusted p-values in square brackets) to account for multiple hypothesis testing in analyzing worker treatment effect heterogeneity shown in Figure 4, based on Westfall & Young (1993) free step-down procedure (5,000 replications) and while accounting for clustering by worker in bootstrapping. The first row's family of hypotheses is four tests on whether the coefficient for Science Info Shock X Characteristics equals zero for the 4 worker characteristics considered in our heterogeneity analysis (quality, gender, STEM degree, and current employment). The second row is analogous to the first row, but for business info shock. The specification is $y_{nf} = \alpha_0 + \alpha_1 \text{BizInfoShock}_n + \alpha_2 \text{SciInfoShock}_n + \alpha_3 \text{C}_n + \alpha_4(\text{BizInfoShock}_n \times \text{C}_n) + \alpha_5(\text{SciInfoShock}_n \times \text{C}_n) + \boldsymbol{X}_{nf} + \varepsilon_{nf}$, where y_{nf} is the key dependent variable of applying to a job.

Table A11: Differences in Worker and Venture Average Characteristics in the 20% Most and Least Affected Observations by Responsiveness to Information Shocks

	Science Info Shock				Business Info Shock			
	Estimate	S.E.	*jp*-value	*p*-value	Estimate	S.E.	*jp*-value	*p*-value
Panel A: Worker characteristics								
>Avg quality	−0.05	0.25	1.00	0.43	1.01	0.25	0.01	0.00
Male worker	0.89	0.20	0.00	0.00	0.69	0.25	0.10	0.00
STEM	−0.09	0.23	1.00	0.34	−0.10	0.33	1.00	0.38
Employed	−0.34	0.22	0.60	0.06	0.03	0.24	1.00	0.45
Panel B: Venture characteristics								
BizDev exp	0.44	0.32	0.54	0.08	−1.17	0.33	0.01	0.00
PhD Founder	−0.51	0.28	0.33	0.03	0.61	0.42	0.52	0.07
Post-revenue	0.88	0.25	0.02	0.00	0.19	0.42	0.98	0.33
Has financing	1.09	0.28	0.01	0.00	0.02	0.43	1.00	0.48

Notes: This table shows the difference in average characteristics of workers (Panel A) and ventures (Panel B) between the 20% most and least affected job applications by science and business information shocks in the Primary RCT. Results are based on the Sorted Effects method of Chernozhukov *et al.* (2018) and is implemented using the R package by Chen *et al.* (2019).

Table A12: Credible Quality Signals in Startup Job Advertisements

Signal	% of All Jobs	% of Business Development Jobs
Founder Education	3.4	2.9
Academic Spinout	1.3	1.0
Other Spinout	0.2	0.2
Incubator Participation	4.8	3.6
Formal IP	2.0	2.4
Named Buyer or Partner	5.6	7.1
International Sales	1.4	1.5
Named Investor or Large Grant	7.5	8.5
Unnamed Investor's Prior Exits	0.2	0.2
Prize or Contest Winner	1.3	1.7
Prominent Advisor	0.2	0.5
Founder's Startup/Corporate Experience	1.8	1.2
Founder's Award for Related Work	0.5	1.0
Media Mention	1.4	1.2
Tech Based on Published Science	0.4	0.2
Specific Sales Traction	0.2	0.2
At least one credible signal	22.7	24.3
Product Description	92.6	94.4
Technical Description	24.6	17.8
Business Model/Monetization Strategy	5.4	9.5

Notes: This table shows characteristics of the universe of job advertisements (N=1017) on AngelList Careers during a two-week period from startups with 1-10 employees. "% of All Jobs" refers to the fraction of job ads which mention each feature. "% Business Development Jobs" restricts to the 411 job ads which are not technical or engineering hires. See Appendix E for the description of the features.

Table A13: Correlation between Success Beliefs and Applications

	(1)	(2)	(3)
Panel A: Dep. Var. = Applied			
Pr(Raise at 1m Valuation)	0.005***		0.006***
	(0.001)		(0.001)
Pr(Successful Exit)		0.002***	−0.000
		(0.001)	(0.001)
R^2	0.14	0.09	0.14
Observations	534	534	534
Panel B: Dep. Var. = Top Rank			
Pr(Raise at 1m Valuation)	0.000		0.000
	(0.000)		(0.000)
Pr(Successful Exit)		0.000	0.000
		(0.000)	(0.000)
R^2	0.03	0.03	0.03
Observations	534	534	534
Panel C: Dep. Var. = Top 3 Choices			
Pr(Raise at 1m Valuation)	0.002***		0.002***
	(0.001)		(0.001)
Pr(Successful Exit)		0.001	−0.000
		(0.001)	(0.001)
R^2	0.07	0.05	0.07
Observations	534	534	534
Panel D: Dep. Var. = Normalized Rank			
Pr(Raise at 1m Valuation)	0.011***		0.011***
	(0.002)		(0.003)
Pr(Successful Exit)		0.005**	−0.000
		(0.002)	(0.002)
R^2	0.13	0.08	0.13
Observations	534	534	534

Notes: This table shows within-startup correlations between worker success beliefs and job applications in the Primary RCT. Beliefs are the incentivized probabilities that the startup will raise external capital at $1m valuation, and experience an IPO or an acquisition with $50m or above valuation. Data is 604 responses from Primary RCT, of which 7 are missing Pr(Raise at 1m Valuation), 4 are missing Pr(Successful Exit), and 59 are missing both. Standard errors clustered by worker in parentheses.

Appendix B Additional Discussion

B.1 Other Related Work in Management and Finance

While our paper primarily contributes to the literature in personnel economics and labor economics, there is also work related to our paper in management and finance. In management, Aran & Murciano-Goroff (2023) conduct a survey experiment with college-educated workers in startups, finding that many exhibit limited financial literacy about the value of startup equity. Focusing on engineers, Tambe *et al.* (2020) show that many workers in information technology place significant value on learning new skills. This suggests that there are other non-pay considerations besides probability of a successful exit that could be important for startup employees. Roach & Sauermann (2023) argue that PhD scientists join startups despite lower wages because ability and preference for startups are uncorrelated, allowing startups to hire high-ability, strong-preference candidates. Beckman & Burton (2008) find that startups who do not hire important functional business roles early on, when they don't have those skills on the founding team, have a lot of trouble hiring those roles as the firm grows. Honoré & Ganco (2022) show that workers avoid startups that are not spinouts (i.e., that do not have obvious pre-existing links to an industry) unless they have a large founding team that serves as a substitute measure of quality. In finance, Bernstein *et al.* (2022) show that job interest in startups rises when a prominent venture capitalist invests. Bernstein *et al.* (2024) show that workers on AngelList became more likely to apply to safer startups during covid. Overall, we view our results as highly consistent with and complementary to these other studies, which also paint a picture of limits to sophistication and significant information frictions for startup employees.

B.2 Discussion on the Quadratic Scoring Rule

We further discuss our system for incentivizing beliefs (i.e., our quadratic scoring rule), expanding further on footnote 11 and Section 5.1 in the main text. One concern with our results on worker beliefs is whether they are driven by our use of a quadratic scoring rule. Danz *et al.* (2022) show in a lab that the binarized scoring rule, which is broadly similar to our risk-invariant quadratic scoring rule of McKelvey & Page (1990), often exhibits measurement error in measuring subject beliefs. If there is classical measurement error in beliefs, this will not lead to bias for our regressions of belief on treatment, nor will it bias our conclusion that workers are overoptimistic about the probability of positive firm events. It will contribute to larger standard errors. A key thing about our use of a quadratic scoring rule is that we explicitly tell workers that it is incentive-compatible to state their true beliefs following work such as Hoffman (2016). Wang (2011) finds that quadratic scoring rules yield more accurate beliefs than non-incentivized beliefs, and Palfrey & Wang (2009) find that quadratic scoring produce more accurate beliefs relative to an improper scoring rule (namely, a simple linear penalty scoring rule), though not all work supports that incentives improve accuracy. For example, Hoffman & Burks (2020) randomize whether workers receive a quadratic scoring rule incentive in guessing about their productivity, and find that the scoring rule has little effect on beliefs. Haaland *et al.* (2023) provide general discussion on measuring beliefs, arguing that belief questions can yield meaningful data even without incentives. In our setting, we believe the quadratic scoring rule incentives serve to draw in job applicants' focus, and that it is highly unlikely the incentives decrease the quality of the belief elicitation.

B.3 Details on RCT Timing/Registration, Scientific Scores, and RSD Procedure

Here, we expand more on timing for the two RCTs, including when they were registered. We also provide details on the science expert scores and the procedure for the random serial dictatorship (RSD) mechanism.

Timing and registration. The primary RCT was conducted during May-August 2019. The RCT was registered with a pre-analysis plan in the AEA RCT Registry in August 2019 before data collection had completed and before data analysis had occurred. The secondary RCT was conducted before this in March 2018. The secondary RCT was used to help select students for entrance into the SEP MBA class, and it was unclear at the time whether the results would be used for research purposes, or whether we could move forward with a broader research study, which also required buy-in from the business schools, so that their alum could be contacted about the job board.

Is it any concern for our paper's conclusions that data from the secondary RCT were analyzed before the RCT was registered? In our view, the answer is strongly no. Our paper's main outcome variables (i.e., job applications and the incentivized firm ranking list) are exclusively from the primary RCT and thus do not face this concern. The paper's key findings are robust to restricting to data from the primary RCT.

Scientific scores. No scientific evaluations were done for 9 of the 26 startups in the primary RCT, generally because their product did not rely on novel science.[1] These firms were considered below-median (or not above-median), keeping with the idea that quality of the underlying science is not a source of competitive advantage for these firms. Nonetheless, our main results on job applications are highly robust to excluding these 9 firms. Figure A1 shows the distribution of both science and business scores. For each job seeker in the primary RCT, the 3 startups randomly chosen for belief questions were selected from the 17 firms for which scientific evaluations were done.

RSD procedure. In the primary RCT, workers are informed that their job applications will be passed along to firms according to the RSD mechanism, where a job application would be passed along to a startup based on their ranking. Once workers had already submitted all their job applications, the actual implementation by SEP was slightly different, though still very much in line with what workers were informed. Firms received a zip folder containing the resumes of all workers who ranked the firm, but firms were provided a short list of applicants whose names were included based on the RSD mechanism. That is, workers ended up receiving slots on the special short list of applicants passed along to the firms, and firms were informed that the slots were allocated based on RSD. That implementation occurred in this manner has no effect on the conclusions or interpretation of the paper. Since

[1] For 8 of the 9 firms, SEP is confident that the firm's product did not rely on novel science. The 9th firm's science quality was uncertain and was a late entry to SEP for idiosyncratic reasons.

the zip folder contained many resumes, being on the short list is akin to have your application forwarded by SEP, with the other applications arriving through another channel.

Appendix C Theory Appendix

A Model of Hiring with Asymmetric Information

In this Appendix, we present a stylized model of hiring under imperfect information about firm quality. We use it to show that providing expert ratings (1) increases the number of applicants who apply to above-average firms, (2) decreases the number who apply to below-average firms, and (3) increases the total surplus generated, and wage inclusive of firm equity, paid by high-quality startups.

There are two fundamental assumptions in the model, both of which match our experimental setting. First, workers do not perfectly learn the quality of startups they apply to until they make a costly application. Second, the nature of this imperfect information is that workers are sometimes unable to tell the difference between more promising and less promising startups, not that workers simply observe firm quality with noise. That is, in expectation, both high-quality and low-quality startups will be seen as being closer to the median firm that they actually are.

The reason the model is game-theoretic (in a very simple way) is to account for job-seekers potentially competing with one another for jobs.

Primitives: Let there be M firms and $N \geq M$ workers. Let the surplus generated from firm j hiring worker i be $\Pi_{ij} = q_{ij}Q_j$, where $Q_j > 0$ are fixed firm qualities and $q_{ij} > 0$ are worker match qualities drawn from an i.i.d. distribution with mean q.[2] That is, the surplus created by a given firm and given worker is weakly complementary in the quality of the other.

Information Asymmetry: Workers do not observe Q_j directly before applying. Rather, all workers observe a common signal μ_j for each firm. For a fraction $\delta \in (0, 1)$ of firms, drawn randomly, $\mu_j = Q_j$, the true firm quality. For the remaining fraction $1 - \delta$ firms, $\mu_j = 0$, an uninformative signal that pools each of these firms. Neither workers nor firms observe their match-quality q_{ij} until worker i applies to firm j.

Timing: First, all workers commonly observe signals μ_j for each firm j. Second, workers apply to exactly one firm; this is a reduced-form equivalent to assuming a linear cost per application of c such that in equilibrium no worker applies to more than one firm. Third, workers and the firms they apply to observe match-specific qualities q_{ij}. Fourth, firms hire the best worker that applied. Finally, any worker who is hired earns payoff equal to a fixed share of surplus $\alpha\Pi_{ij}, \alpha \in (0, 1)$; that is, workers are given an equity share in the firm they work for. Note that due to the surplus sharing assumption in this model, policies that

[2] With some algebraic complexity, this model can be extended to handle workers with heterogeneous quality. High-quality workers are equally dissuaded from applying to the best firms due to information asymmetry as low-quality workers: both have imperfect information about the true quality of firms.

maximize worker payoff, firm surplus, and total surplus are identical.

Let us now solve the model, denoting with p_{ij} the probability worker i applies to firm j. Since the share δ of firms with an uninformative signal are chosen at random, workers' posterior belief of the quality of these firms will be exactly $\bar{Q}$, the average quality of the firms whose quality is observed. Let $\bar{\mu_j} = \bar{Q}$ for firms with these uninformative signals, and $\bar{\mu_j} = \mu_j = Q_j$ for all other firms.

Workers will maximize their payoff from a given match times the probability they are hired. Since workers are identical other than their idiosyncratic match quality, the probability a worker gets hired is just the probability their idiosyncratic match-quality is highest, or one over the number of other applicants to the same firm. Therefore, worker i chooses the randomization strategy across firms p_i to maximize the expectation

$$\mathbb{E}[\sum_j \frac{p_{ij}\alpha q_{ij}\bar{\mu_j}}{\sum_i p_{ij}}]$$

We now solve for the symmetric mixed-strategy equilibrium.

Lemma 1 *Assume that there exists a symmetric mixed-strategy equilibrium where workers apply to all firms with positive probability.*[3] *Then:*

1. *In any symmetric mixed-strategy equilibrium, the probability each worker applies to firm j is $p_j = \frac{\bar{\mu_j}}{\sum_{j'}\bar{\mu_j}}$.*

2. *Therefore, the number of applicants for firm j is a binomial distribution with probability $\frac{\bar{\mu_j}}{\sum_{j'}\bar{\mu_j}}$ and N trials.*

Proof: In any mixed-strategy equilibrium, the payoff of applying to firms j and j' in the support must be identical. That is, $\mathbb{E}[\frac{p_{ij}\alpha q_{ij}\bar{\mu_j}}{\sum_i p_{ij}}] = \mathbb{E}[\frac{p_{ij'}\alpha q_{ij'}\bar{\mu_{j'}}}{\sum_i p_{ij'}}]$. Since $\mathbb{E}[\epsilon_{ij}] = 0, \forall i, j$, that equality reduces to $\mathbb{E}[\frac{p_{ij}\alpha q\bar{\mu_j}}{\sum_i p_{ij}}] = \mathbb{E}[\frac{p_{ij'}\alpha q\bar{\mu_{j'}}}{\sum_i p_{ij'}}]$. Hence for any firms j and j', $\frac{p_{ij}}{p_{ij'}} = \frac{\mu_j}{\mu_j'}$, and by symmetry, $\frac{p_j}{p_{j'}} = \frac{\mu_j}{\mu_j'}$. Summing this equality for all $j' \neq j$, we have that $p_j = \frac{\bar{\mu_j}}{\sum_{j'}\bar{\mu_j}}, \forall j$. The second part of the lemma follows immediately. ■

The previous lemma says that the expected number of applicants to a given firm is increasing in the workers' posterior belief $\bar{\mu_j}$ of the firm's quality.

Proposition 2 *Let an information treatment increase δ, the probability workers observe true firm quality.*

1. *Above-average firms receive more applications.*

2. *Below-average firms receive fewer applications.*

[3]This requires that the worst firm is not so bad that workers would avoid applying even if they were guaranteed a job at that firm as the only applicant. That is, $\min_j Q_j$ needs to be sufficiently high.

3. *The change in the number of applications a firm receives when workers gain perfect information about the firm's quality is increasing in the difference between the firm's true quality and the average quality of all other firms.*

4. *Surplus generated by above-average firms, and hence wages for workers they hire, increases.*

Proof: by the previous lemma, the number of workers that apply to firm j in expectation is increasing in the perceived quality of the firm $\bar{\mu}_j$. When workers do not perfectly observe the quality of firm j, in expectation workers believe that firm to have equal quality to the average of all other firms.[4] Therefore, the expected number of applicants to firm j is

$$\delta \frac{Q_j}{\mathbb{E}[\sum_{j' \neq j} Q_{j'}]} + (1-\delta) \frac{\mathbb{E}[Q_{j' \neq j}]}{\mathbb{E}[\sum_{j' \neq j} Q_{j'}]}$$

Therefore, for firms with $Q_j > \mathbb{E}[Q_{j' \neq j}]$, an increase in δ raises the probability workers believe the firm to have a higher quality, and hence raises the expected number of applicants. Likewise, for below-average firms where $Q_j < \mathbb{E}[Q_{j' \neq j}]$, an increase in δ decreases the expected number of applicants. Finally, if K_j workers apply for firm j and the firm hires the best worker who applies inclusive of idiosyncratic match quality, total expected surplus is $X(K_j)Q_j$ where $X(K_j)$ is the $K_j th$ order statistic from the distribution q_{ij} is being drawn from. That is, $X(K_j)$ is the expected quality of the best applicant who applies to firm j conditional on getting K_j applications. An increase in the expected number of applicants therefore also increases expected surplus earned by a given firm. ■

The proposition above is not simply the result of asymmetric information about firm quality. For instance, if workers received a signal with mean-zero noise about each firm, that noise could both *increase* or decrease the number of applicants a firm gets: sometimes the noise causes workers to overestimate the quality of even the best firms. The fundamental issue in our empirical setting is not the misperception, but the *pooling* of high- and low-quality firms. Information increases applications when it affects the relative quality of my firm relative to others.

[4] That is, when firm j has its true quality hidden, $\mathbb{E}[\bar{Q}]$ across all realizations of firms that could have their true quality hidden from workers is just the expected true quality of all firms other than the focal firm.

Appendix D Screenshots from the Primary RCT

Figure D1: Primary RCT Survey Instructions

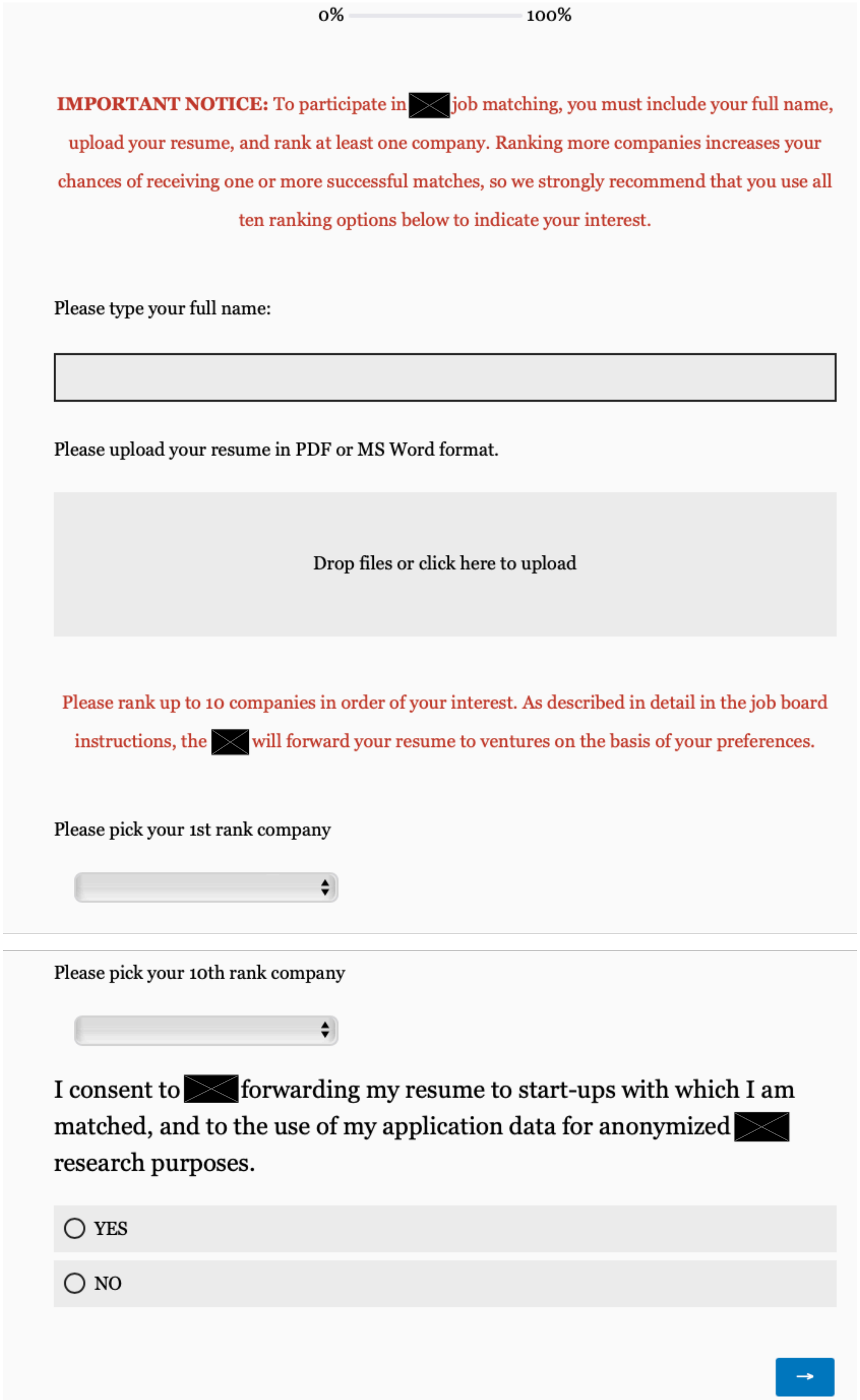
Instructions

Welcome to the inaugural [redacted] job board. [redacted] is the largest science-based start-up mentorship program in the world. 26 start-ups in this year's cohort expressed interest in talking to [redacted] alumni interested in business development jobs. Through this system, we will forward your resume and information to start-ups of interest.

Clicking on company logos brings up brief descriptions written by each venture. After examining these companies, upload your resume in the form to the right, and list the 10 companies of greatest interest in your order of preference from most preferred ("1") on down. **If you do not see a form on the right side of this page to rank companies, please turn off your ad blocker or try with a different browser**.

To avoid inundating these start-ups with an excessive number of resumes, we have agreed to forward a limited number of resumes to each start-up. **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.

Figure D2: Screenshots from the Primary RCT (highlighting from the original)



0% 100%

IMPORTANT NOTICE: To participate in [redacted] job matching, you must include your full name, upload your resume, and rank at least one company. Ranking more companies increases your chances of receiving one or more successful matches, so we strongly recommend that you use all ten ranking options below to indicate your interest.

Please type your full name:

Please upload your resume in PDF or MS Word format.

Drop files or click here to upload

Please rank up to 10 companies in order of your interest. As described in detail in the job board instructions, the [redacted] will forward your resume to ventures on the basis of your preferences.

Please pick your 1st rank company

Please pick your 10th rank company

I consent to [redacted] forwarding my resume to start-ups with which I am matched, and to the use of my application data for anonymized [redacted] research purposes.

YES

NO

→

0% 100%

Your response has been recorded.

The next 5 questions are optional and will not affect job matching. However, if you complete them, you may win up to $250 and will inform [redacted] design of its Job Board.

What is your assessment of the **quality of the science/technology** in the below ventures, on a scale from 1 (lowest score) to 5 (highest score)?

NOTE: In this setting, the quality of the science is defined as your overall assessment of the quality of the underlying science and its potential for forming the basis of a commercial product.

	(lowest score) 1	2	3	4	(highest score) 5
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○

← →

0% 100%

What is your assessment of the **quality of the business model** for these company, on a scale of 1 (lowest score) to 5 (highest score)?

NOTE: In this setting, the business model may be defined broadly, as the overall quality and execution potential of the company's business strategy in building a scalable technology-based start-up.

	(lowest score) 1	2	3	4	(highest score) 5
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○

← →

0% 100%

PROBABILITY:

The below questions ask you to think about the **percent chance** that something will happen in the future.

The **percent chance** can be thought of as the number of chances out of 100. You can use any number between 0 and 100.

For example, numbers like:
2 and 5 percent may be "almost no chance",
20 percent or so may mean "not much chance",
a 45 or 55 percent chance may be a "pretty even chance",
80 percent or so may mean a "very good chance",
and a 95 or 98 percent chance may be "almost certain"

INCENTIVES ON PROBABILITY QUESTIONS:

As added encouragement on probability questions, five people who complete this survey will be chosen at random to be paid via a lottery system. For those chosen, payment will be $0 or $250. Payment will be based on one of two questions below. This lottery system has been used to elicit people's probability beliefs in various contexts and is specially designed so that **it's mathematically optimal for you to state your true belief about the probability an event will occur.**

For further detail, please see here.

What is the probability that the firms below raise money at a valuation of at least CAD$1,000,000 within 1 year of the time this information was prepared?

NOTE: A $1,000,000 valuation is historically the absolute minimum valuation for a firm which raises a "seed" financing round. Seed financing is generally the first financing round with institutional rather than angel investors. The firm you are evaluating was randomly selected from a sample where historically similar firms have a 35% chance of reaching a $1,000,000 valuation.

Almost No Chance		Not Much Chance			Pretty Even Chance		Very Good Chance		Almost Certain	
0	10	20	30	40	50	60	70	80	90	100

[redacted]

[redacted]

[redacted]

What is the probability that the firm in question has an initial public offering (IPO) or is acquired at CAD$50,000,000 or more within 1 year of the time the information was prepared?

Almost No Chance		Pretty Even Chance			Very Good Chance		Almost Certain		Not Much Chance	
0	10	20	30	40	50	60	70	80	90	100

[redacted]

[redacted]

[redacted]

← →

0% 100%

On a scale from 1 (least interested) to 5 (most interested), how interested would you be in working at this firm?

NOTE: This question will not be used to match you to companies, which is based on your rankings; we are using this question for informational purposes only.

	(lowest score) 1	2	3	4	(highest score) 5
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○
[redacted]	○	○	○	○	○

← →

Appendix E Data Appendix

E.1 Creating Our Samples of Workers and Firms

Candidates contacted. For the Primary RCT, we contacted business school alumni from two North American business schools in three different email campaigns. From the first business school, we reached out to 7,894 Master's alumni on May 27, 2019 with a reminder sent on June 21, to 5,721 undergraduate alumni from the same school on June 25, 2019 with a reminder sent on July 9. For the second business school, we reached out to 3,701 undergraduate and 2,083 Master's alumni on August 6, 2019, though the school preferred not to send a second reminder. Furthermore, 40 double-degree alumni from the first school were contacted twice. In calculating our balance tables, we remove these individuals from the undergraduate subsample, resulting in 5,681 undergraduate alumni. Thus, we contacted 19,359 unique alumni. PhD alum were not emailed.

Subjects. We received 264 survey submissions. We dropped from this set 4 repeat submissions from identical individuals where we only kept their earlier time-stamped submission. We dropped 1 dummy submission by a staff member of the SEP. Finally, we dropped 9 submissions by individuals who we could not match to our original contact lists. We managed to contact 5 of these individuals; they all told us that their friends and family had forwarded the Job Board email invitation to them. The remaining 250 submissions constitute our Primary RCT data.

In Section 3 of the main text, we discuss the tracking data which allow us to provide information on the participation rate conditional on clicking on the job board link. As mentioned, there were 587 recipients who opened the email and clicked on the job board link. 93 out of 587 recipients clicked the link in the reminder email. It is possible that some of those 93 people overlap with the initial 494 people who clicked from the original email. This makes the lower bound of unique individuals equal to 494, which would make the participation rate even higher than 46%. Within each individual email, people are unique.

Startups. Primary RCT ventures are 26 startups recruited from the SEP's 2018-19 cohort. The Secondary (pilot) RCT ventures are 26 startups selected from the 2017-18 cohort. The analysis in Table 1, which shows the correlation between expert ratings and performance, uses the full cohort of 130 startups in 2017-18, from which 24 startups were dropped due to missing one or both scores.

E.2 Random Assignment using Unique URLs

It would not be possible for a job seeker to know from the website link that an experiment was going on. The randomization trigger was part of a non-descript block of text such as "nytimes.com/sports?id=Aa674k". The text following the question mark is referred to as a "url query". In nine cases, dropped from our analysis, we received applications from people who were not the original targeted recipient of the job board invite.

E.3 Description of Variables

Gender. Obtained by linking first names to the US Social Security Administration's list of most common names by gender. Names with a greater than 50% chance of being male are classified as male, whereas names with a greater than 50% chance of being female are classified as female.

Race. By clicking on the link provided on the resume, race is obtained by checking the social media (typically LinkedIn) profile of the candidate. In the absence of a web link, individual names are searched online and identification is ensured by cross referencing profile information with resume information. All other cases are obtained by linking last names with census information on the distribution of race by last name.

City is SEP HQ is an indicator equal to 1 if worker's lives in the same city as SEP HQ.

Graduation year is a worker's year of graduation obtained from the business school registrar's office records.

Startup founder is an indicator from resumes equal to 1 if the applicant founded a business.

Startup employee is an indicator from resumes equal to 1 if the worker has startup employment experience.

Employed is an indicator from resumes equal to 1 if the worker is currently employed.

Years of experience is an integer from resumes equal to the total years of worker experience rounded to the nearest year. Includes internships.

STEM is a binary variable from the applicant's resume that is equal to 1 if the candidate listed an undergraduate degree in natural, formal, or engineering sciences.

Worker Quality (1-10) is a number from 1 to 10 that reflects candidate quality for a business development job at a fast-growing, science-based startup that has just received early venture capital investment (10 is the highest score). An independent startup-focused HR expert determined these scores based on de-identified worker resumes.

Predicted Salary (Thousand) is the annual salary the worker should be offered in order for the startup to have a chance at hiring them. An independent experienced HR consultant determined these salaries based on de-identified worker resumes.

Number of founders is an integer that is equal to the size of the founding team reported on the application form of the firms submitted to SEP in the summer of 2018.

Number of employees is an integer that is equal to the number of non-founding employees reported on the application form of the firms submitted to SEP in the summer of 2018.

PhD founder is an indicator from startup applications to SEP equal to 1 if the startup had at least one PhD founder.

Technology fixed effects are a series of dummy variables reflecting the core technology of the startup. These include machine learning, quantum machine learning, blockchain, space, cities, and health.

Pct SEP Activities Completed is the fraction of high-priority business objectives firms completed during SEP. Every eight weeks, founders and mentors set three objectives that constitute the highest priorities of startups. SEP then verifies whether each objective is completed.

BizDev experience is an indicator from startup applications to SEP equal to 1 if the startup had at least one founder with business development professional experience.It includes experience in marketing, operations, finance, or other executive roles.

Top 1/3 Page is an indicator equal to 1 if the startup's profile is positioned in the top one-third of the website.

Raised capital is an indicator obtained from SEP internal data equal to 1 if the firm raised external capital before the experiment.

E.4 Description of AngelList classification

We examine the full text of all 1017 advertisements for a full-time job posted on AngelList's job board between October 30 and November 13, 2020 by a startup with between 1 and 10 employees. 40.4% of the advertisements are for non-technical roles, including sales, marketing, upper management, HR, communications, and finance. The remainder are technical roles, largely engineering. From each advertisement, we hand-code the following variables.

Founder Education. The advertisement lists the university at least one founder holds a degree from.

Academic Spinout. The advertisement describes the firm as based on, or a spinout from, an academic lab or academic research performed by the founding team.

Other spinout. The advertisement describes the firm as based on, or a spinout from, work done at an incumbent firm or government agency.

Incubator Participation. The advertisement lists participation in a named incubator or accelerator.

Formal IP. The advertisement notes that the firm holds formal IP such as a patent or pending patent.

Named Buyer or Partner. The advertisement specifically names a current customer or partner.

International Sales. The advertisement notes the company has made sales beyond its country of origin.

Named Investor or Large Grant. The advertisement notes that the company has received funding from a named investor, foundation, or government agency.

Unnamed Investor's Prior Exits. The advertisement describes the company as receiving investment from the backer of a prior named successful startup.

Prize or Contest Winner. The advertisement describes the company as a winner (including non-first prize winners) of any business model, technical, or product contest.

Prominent Advisor. The advertisement describes the firm as being advised or mentored by a specific named person.

Founder's Startup/Corporate Experience. The advertisement describes the founders as having previously led a successful exit, founded a named startup, or worked in an executive position at a related incumbent firm.

Founder's Award for Related Work. The advertisement notes that a founder has won a prize, or is well-known for, work related to the startup.

Media Mention. The advertisement listed a named media source as having written up the company, or the company has participated in a popular entrepreneurship program like Dragon's Den or Shark Tank, or the company has appeared on Product Hunt.

Tech Based on Published Science. The firm's technology is described as being derived from published, peer-reviewed scientific work.

Specific Sales Traction. Specific sales success, such as a high position on an App Store, are included in the advertisement.

Product Description. The advertisement describes the company's primary product. In general, "stealth" startups are the only ones who do not give this detail.

Technical Description. The advertisement gives specific technical details about the operation or production or nature of the product (e.g., "We use a generative adversarial network to investigate financial fraud...").

Business Model/Monetization Strategy. The advertisement specifically describes how the product is being monetized or scaled (e.g., "We operate a two-sided platform where we charge banks to connect to our high net-worth investors...").

Appendix F Key Documents from the Secondary RCT

The two figures below show a sample startup dossier and treatment shown to MBA students in the Secondary RCT. The top panel shows the background information provided, while the bottom panel shows how expert ratings were displayed for the group that received both the Business Model and Science scores.

[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.

Treatment Shown to the Group Receiving both Business Model and Science Scores

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

Appendix G Survey of Economist Experts

Of the 270 NBER economists contacted, 120 were attendees from NBER Personnel, 100 were attendees from NBER Entrepreneurship, and 50 were attendees from NBER Labor.[5] Our 32% response rate from NBER economists is slightly higher than that in a leading recent study by Deshpande & Dizon-Ross (2023) who receive a response rate of 24% in surveying members from the NBER Children and Education groups. Note that we do not know the job titles (e.g., full professor, associate professor, ...) for the 10 responses from the Social Science Prediction Platform. Results in the economist survey are qualitatively similar when restricting to faculty members.

As in any expert prediction exercise, it is critical that experts are not already familiar with the results of the study. We addressed this point in two ways. First, in drawing our base survey sample of 270 NBER economists, we manually excluded several economists we believed were familiar with the results (e.g., by seeing the paper at a seminar). Second, as described in the main text, we began the survey by asking a screener question.

Question 1: As a screening question, are you familiar with the main findings from the NBER working paper "Information Frictions and Employee Sorting between Startups?" For example, have you read the paper or its abstract?

- Yes
- No

If the respondent answers "Yes" to being aware of the study's main results, the survey terminates. If the answer is "No," then the respondent sees an overview of the study before proceeding to the prediction questions. Below is the description that respondents saw about our study:

[5] Some economists attend multiple meetings from our set of Personnel, Entrepreneurship, and Labor. We drew first from Personnel attendees, second from Entrepreneurship, and third from Labor. Thus, an economist who attended Personnel and Entrepreneurship would count as a Personnel attendee, and an economist who attended Entrepreneurship and Labor would count as an Entrepreneurship attendee.

Overview of the study

Alum of two North American business schools were invited to participate in a startup job board. This job board featured 26 early-stage science-based startups who had 1) participated in a world-leading entrepreneurship program, and 2) chose to be on the job board. To fix ideas, a typical startup in our setting would be one founded by two computer science professors with an advancement in artificial intelligence for autonomous vehicles.

Treatments

Each job seeker was randomly selected to receive a customized link to one of the four versions of the job board. The **control group** saw a job board with just the ads written by the startups. The remaining three groups saw those ads alongside a note indicating whether the startup's **science** and/or **business** quality received an above-average expert rating.

The science expert rating was determined by a PhD scientist from Canada's National Research Council with expertise in the startup's technological domain. This rating was based on a 30-minute interview with founders and detailed written materials the venture provided in advance of the meeting. The business expert rating was provided by experts with experience in the evaluation of technology-based startups who were asked to evaluate the startups' business models, scalability, and potential to execute. The business rating was based on an in-person interview and extensive documents, as well as informal interactions.

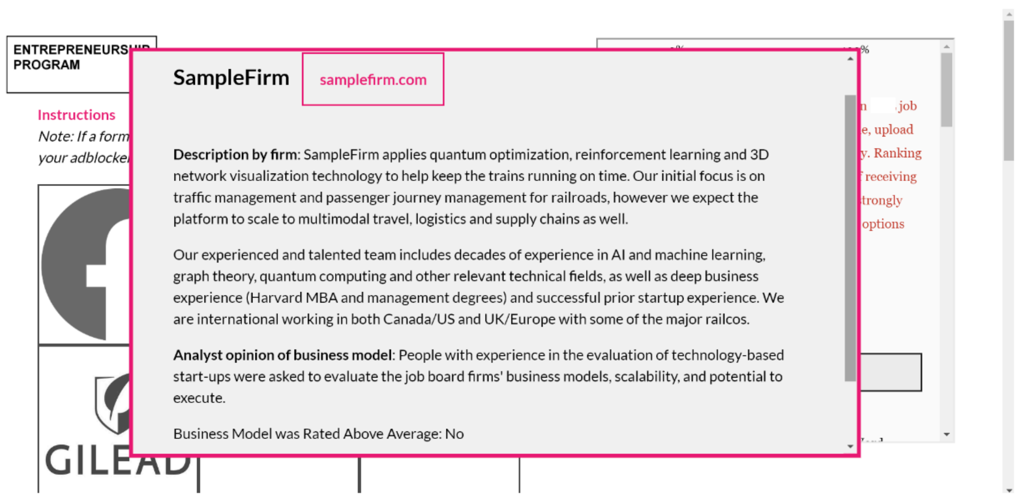
The image below summarizes our 2X2 RCT design:

	Hide Business Rating	**Show Business Rating**
Hide Science Rating	**Standard Job Ad**	**Standard Job Ad + Business Rating**
Show Science Rating	**Standard Job Ad + Science Rating**	**Standard Job Ad + Science & Business Rating**

Note: Standard Job Ad includes the firm's description, names of the founders, and link to the firm's website.

Overview of the study

The below screenshot shows a sample anonymized job ad with a negative business rating. Note that the expert rating treatment was at the **market level** and not at the job ad level--job seekers in a given job board either saw or did not see a given expert rating treatment for all job ads.



Next, we provide the full list of questions in the survey following the screener.

Question 2: Which had a larger effect on job applications, science or business expert ratings?

- Science expert ratings
- Business expert ratings
- Science and business expert ratings had about the same effect
- Both had no effect

Question 3: Which had a larger effect in terms of magnitude on job applications, positive or negative information?

- Positive information (i.e., information that firm quality is above-median within our sample) has a larger effect
- Negative information (i.e., information that firm quality is below-median within our sample) has a larger effect
- Positive and negative information had about the same effect in terms of magnitude
- Both had no effect

Question 4: Were science ratings and business ratings complements or substitutes in terms of their impact on job applications?

- Complements
- Substitutes
- Both had no effect

For the next questions, we define a good firm as one rated above-average in terms of both its science and business by experts. We define a bad firm as one rated below-average in terms of both its science and business by experts.

Please use a response of "X" to predict that good firms received X% more applications than bad firms, and use a response of "-X" to predict that bad firms received X% more applications than good firms. X is the number that you provide.

Question 5: Baseline: In the control group where we showed no expert ratings, how many percent (%) more or less applications did good firms receive compared to bad firms?

Question 6: Impact of Science Expert Ratings: When jobseekers viewed expert science ratings, how many percent (%) more or less applications did good firms receive compared to bad firms?

Question 7: Impact of Business Expert Ratings: When jobseekers viewed expert business ratings, how many percent (%) more or less applications did good firms receive compared to bad firms?

Question 8: Impact of Science and Business Expert Ratings: When jobseekers viewed both expert science and business ratings, how many percent (%) more or less applications did good firms receive compared to bad firms?

Question 9: Do you think the effect of expert ratings on job applications varied based on whether the job seeker had an undergraduate STEM degree?

- Yes, significantly smaller effect for those with a STEM degree than those without
- No difference based on worker STEM background
- Yes, significantly larger effect for those with a STEM degree than those without
- Expert ratings had no effect

Question 10: Do you think the effect of expert ratings on job applications varied based on the "quality" of the job seeker? To measure the job seeker quality, the project partnered with an HR expert who focuses on startup hiring. This HR expert rated the resumes of the job seekers based on suitability for working at a startup, and we divided these into above-median quality and below-median quality.

- Yes, significantly smaller effect for above-median quality candidates than for below-median
- No difference based on worker quality
- Yes, significantly larger effect for above-median quality candidates than for below-median
- Expert ratings had no effect

Question 11: Do you think the effect of expert ratings on job applications varied based on the gender of job seeker?

- Yes, significantly smaller effect for women than men
- No difference based on worker gender
- Yes, significantly larger effect for women than men
- Expert ratings had no effect

Appendix H Detailed Explanation of the Quadratic Scoring Rule

The figure below displays the more detailed explanation of the risk-invariant quadratic scoring rule that was made available to subjects (in addition to the simpler and intuitive explanation that was also provided to subjects, and that can be seen in Appendix D). This explanation was used in both the primary and secondary RCTs. The quadratic scoring rule is used to provide incentives on the probability questions. In the primary RCT, the more detailed explanation was accessed by clicking a link "For Further Detail". In the secondary RCT, the more detailed explanation was given on a separate sheet of paper.

Quadratic Scoring Rule Explanation Sheet

INCENTIVES ON SOME PROBABILITY QUESTIONS:

As added encouragement on probability questions, five people who complete this survey will be chosen at random to be paid via a lottery system. Payment will be based on one of two questions below. This lottery system has been used to elicit people's probability beliefs in various contexts, and is specially designed so that **it's mathematically optimal for you to state your true belief about the probability an event will occur.**

Specifically, if you are randomly chosen for possible payment, you will receive $250 CAD or $0. The probability of receiving $250 is equal to $2p-p^2$ if the event occurs, and is equal to $1-p^2$ if the event does not occur, where p is the probability that you give. The below table gives examples of your probability of winning $250 depending on the probability that you state and whether the event in question occurs or not.

Your Stated Probabilty	Your probability of winning $250 if event occurs	Your probability of winning $250 if event does not occur
0	0.0000	1.0000
5	0.0975	0.9975
10	0.1900	0.9900
15	0.2775	0.9775
20	0.3600	0.9600
25	0.4375	0.9375
30	0.5100	0.9100
35	0.5775	0.8775
40	0.6400	0.8400
45	0.6975	0.7975
50	0.7500	0.7500
55	0.7975	0.6975
60	0.8400	0.6400
65	0.8775	0.5775
70	0.9100	0.5100
75	0.9375	0.4375
80	0.9600	0.3600
85	0.9775	0.2775
90	0.9900	0.1900
95	0.9975	0.0975
100	1.0000	0.0000

Appendix References

Aran, Yifat, & Murciano-Goroff, Raviv. 2023. Equity Illusions. *Journal of Law, Economics, and Organization*, Forthcoming.

Beckman, Christine M., & Burton, M. Diane. 2008. Founding the Future: Path Dependence in the Evolution of Top Management Teams from Founding to IPO. *Organization Science*, **19**(1), 3–24.

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.

Bernstein, Shai, Townsend, Richard R., & Xu, Ting. 2024. Flight to Safety: How Economic Downturns Affect Talent Flows to Startups. *Review of Financial Studies*, **37**(3), 837–881.

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

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

Danz, David, Vesterlund, Lise, & Wilson, Alistair J. 2022. Belief elicitation and behavioral incentive compatibility. *American Economic Review*, **112**(9), 2851–2883.

Deshpande, Manasi, & Dizon-Ross, Rebecca. 2023. The (Lack of) Anticipatory Effects of the Social Safety Net on Human Capital Investment. *American Economic Review*, **113**(12), 3129–72.

Haaland, Ingar, Roth, Christopher, & Wohlfart, Johannes. 2023. Designing Information Provision Experiments. *Journal of Economic Literature*, **61**(1), 3–40.

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

Hoffman, Mitchell, & Burks, Stephen V. 2020. Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits. *Quantitative Economics*, **11**(1), 315–348.

Honoré, Florence, & Ganco, Martin. 2022. Entrepreneurial teams' acquisition of talent: Evidence from technology manufacturing industries using a two-sided approach. *Strategic Management Journal*, **n/a**(n/a).

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

Palfrey, Thomas R., & Wang, Stephanie W. 2009. On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization*, **71**(2), 98–109.

Roach, Michael, & Sauermann, Henry. 2023. Can Early-Stage Startups Hire Talented Scientists and Engineers? Ability, Preferences, and Employee Job Choice. *Management Science*, Forthcoming.

Tambe, Prasanna, Ye, Xuan, & Cappelli, Peter. 2020. Paying to Program? Engineering Brand and High-tech Wages. *Management Science*, **66**(7), 3010–3028.

Wang, Stephanie W. 2011. Incentive effects: The case of belief elicitation from individuals in groups. *Economics Letters*, **111**(1), 30–33.