The Selection of Talent
Experimental and Structural Evidence from Ethiopia*

Girum Abebe[†], Stefano Caria[‡], Esteban Ortiz-Ospina[§]

July 28, 2020

Abstract

We study how search frictions in the labour market affect firms' ability to recruit talented workers. In a field experiment in Ethiopia, we show that an employer can attract more talented applicants by offering a small monetary incentive for making a job application. Estimates from a structural model suggest that the intervention is effective because the cost of making a job application is large, and positively correlated with jobseeker ability. We provide evidence that this positive correlation is driven by dynamic selection. In a second experiment, we show that local recruiters underestimate the positive impacts of application incentives.

*We are grateful to Abi Adams, Arun Advani, Nava Ashraf, Oriana Bandiera, Vittorio Bassi, Gaurav Chiplunkar, Stefano DellaVigna, David Eil, Marcel Fafchamps, Erica Field, James Fenske, Simon Franklin, Douglas Gollin, Clément Imbert, Supreet Kaur, Philipp Kircher, Rocco Macchiavello, Jeremy Magruder, David McKenzie, Guy Michaels, Paul Niehaus, Barbara Petrongolo, Simon Quinn, Imran Rasul, Chris Roth, Yared Seid, Alemayehu Seyoum Taffesse, Chris Woodruff, Liam Wren-Lewis, Yanos Zylberberg and seminar audiences at NBER Summer Institute, BREAD/CEPR/STICERD/TCD (LSE), Warwick, Toulouse, Cornell Dyson, Sciences Po, PSE, McGill, IFS, NEUDC (Tufts), Mannheim, Oxford, Bristol, Essex, Royal Holloway, Cologne, Ohlstadt for helpful comments. We thank Ian Jewitt for invaluable discussion of the theoretical model. Koen Maskaant and Alemayehu Woldu provided outstanding research assistance. The project would not have been possible without the support of Rose Page, Simon Quinn, CSAE and EDRI. We acknowledge funding from the IGC (Project No. 1-VCC-VETH-VXXXX-32304) and the John Fell Fund. The trial was registered on the AEA RCT registry (AEARCTR-0001900).

[†]World Bank. Email: gtefera@worldbank.org.

[‡]University of Warwick. stefano.caria@warwick.ac.uk, Web: www.stefanocaria.com.

[§]Our World in Data and University of Oxford. Email: esteban@ourworldindata.org.

1 Introduction

Selection problems occur everywhere in society. Firms want to hire talented workers; universities need to attract high-ability students; welfare programs have to select poor recipients. A key insight that often guides the design of selection policies is that a costly application process can improve selection by discouraging the participation of unwanted candidates. This logic underpins ordeal mechanisms in welfare programs (Nichols and Zeckhauser, 1982; Alatas et al., 2016) and is often applied in recruitment and marketing (Ashraf et al., 2010; Bandiera et al., 2011; Alonso, 2018). In this paper, we provide the first experimental evidence showing that the opposite is true in an important economic context — *decreasing* application costs significantly improves selection. Our results are driven by the fact that high quality candidates face on average higher application costs. We quantify this positive correlation between applicant quality and application costs, provide evidence on the mechanism that generates it, and show how it can be leveraged for policy.

Our evidence comes from studying an employer who wants to attract talented workers for a clerical position in Addis Ababa, Ethiopia. There are two important features in this context. First, job search and application costs are high on average; formal jobs require an application in person, which is time consuming and often requires the use of public transport. Second, these costs are heterogenous, as individuals have different access to liquidity and live at a varying distance from employers.

Using unique high-frequency data on job search and employment, we document that this heterogeneity generates a dynamic selection effect. High-ability individuals who face relatively low application costs find work faster and stop searching for work earlier than individuals who have similar ability, but face higher application costs. Thus, over time, a positive correlation between ability and application costs emerges among individuals searching for work.[1] This correlation overturns the standard intuition on the screening role of application costs and suggests that lowering costs may actually be a beneficial policy for an employer in this market. We test this prediction empirically with our experiment.

In the experiment, the employer reduces application costs by offering a small monetary payment to all job applicants. This monetary incentive is worth 4.5 USD and is calibrated to reimburse applicants for both transport costs — an in-person application is required for this position — and the opportunity cost of time. In a second treatment, the employer doubles the wage offer but does not provide any financial incentive for

[1] At the end of the paper, we present evidence from a number of studies suggesting that a similar correlation may be found in other contexts as well, in both developed and developing countries.

applications. The expected value of the wage increase is 105 USD. The employer randomises the offer of these two treatments over the sample of individuals who call to inquire about the position.

Our key finding is that the application incentive *improves* the quality of the applicant pool. We measure applicant ability through standardised tests of cognitive and non-cognitive ability, job-specific experience, and GPA (a proxy of ability widely used in this context).[2] Applicants from the application incentive group have higher cognitive ability and GPA compared to control applicants, and similar levels of non-cognitive ability and experience. The ability gains occur across the whole distribution of ability and are particularly large at the top. GPA and cognitive ability among the highest-scoring applicants invited for an interview increase by .5 and .3 standard deviations, respectively. The number of top applicants (defined as those with cognitive ability above the 90th percentile of the control group distribution) doubles. We also find that raising the wage increases average applicant ability. The magnitude of this effect is similar to the impact of the application incentive, although the two interventions operate through different mechanisms, as we discuss in detail below.

These findings surprise local employers. To show this, we sample 196 recruiters in the same market and ask them to forecast the effect of the application incentive, after being informed of the characteristics of the applicant pool in the control and high wage conditions (DellaVigna and Pope, 2016). The majority of employers underestimate the effect of application incentives and the average employer incorrectly expects this intervention to decrease applicant quality. However, when asked to invite selected jobseekers to make an application at their firm on the basis of anonymised CVs, employers strongly prefer the applicants from the incentive treatment over the applicants from the other two experimental groups. This gives us further evidence that application incentives attract applicants that are preferred by employers.

The improvement in applicant ability generated by the incentive treatment is driven by women, and by those jobseekers who are currently unemployed and less-experienced — all groups that do not usually perform well in the labour market. This is not the case for the high wage treatment. This suggests that the application incentive does not increase ability at the cost of attracting individuals who have better outside options and, hence, a lower likelihood of accepting and keeping the job. On the contrary, this intervention mostly taps the pool of low-income jobseekers who stand to benefit the most

[2] We use the Raven and Stroop tests for cognitive ability (Schmidt and Hunter, 1998). For non-cognitive skills we administer the Big-5 personality test and the Grit scale (John and Srivastava, 1999; Duckworth et al., 2007). To identify job-specific experience, we follow Autor and Handel (2013) and collect measures of experience in several relevant tasks.

from the job. To explore this point further, we generate an individual measure of the net present value of the experiment's job using a simple calibration framework and a Post-LASSO forecast of each individual's market wage (Belloni et al., 2014). We find that the increase in ability is significantly larger for the group of jobseekers that values the job the most.

We rule out several potential explanations for our findings that are not related to job application costs. First, the interventions could induce individuals to exert greater test effort. To study this, we administer a task that requires effort, but very little ability. We do not find significant differences in performance on this task, which provides evidence against differential test effort. The positive impacts on GPA, a measure of ability determined before the interventions, further support this conclusion. Second, the application incentive could help individuals overcome self-control problems. However, we find that incentive group applicants are as likely to be present biased as control group applicants, which is inconsistent with this explanation. Third, subjects could misinterpret the offer of the incentive as a signal about their quality or about the tightness of the labour market. Contrary to this hypothesis, we show that the incentive does not affect subjects' expectations about how long it will take them to find a new job, or the wage that this new job would pay. We also find that the incentive is associated with only minor changes in beliefs about the attributes of the job. We estimate that these changes can account for only 5 percent of the total effect of the application incentive.

To shed light on the structural features of the labour market that drive our reduced-form results we propose a simple model of application decisions. The model captures two key frictions in job search: application costs and uncertainty about the probability of being offered the job. There is only one type of vacancy, but we allow for worker heterogeneity — jobseekers differ in terms of their ability, the magnitude of the application costs they face, and the benefit that they derive from being offered the job. Using this model, we show formally that the incentive can attract better applicants in markets where higher-ability jobseekers face larger application costs.

We identify and structurally estimate the key parameters of the model using the exogenous variation generated by the experiment (DellaVigna, 2018). The fit between the simulated and empirical moments is good. We fit all application rates with less than one percentage point of error. Further, the model can match a key non-targeted moment — jobseekers' assessment of the probability of receiving a job offer — and replicates non-targeted patterns of the data such as the fact that control applicants are negatively selected. Our structural and observational evidence shows that jobseekers are substantially overconfident about their labour market prospects, a finding that is consistent with the evidence reported in Spinnewijn (2015) and Banerjee and Sequeira (2020) for the US and South Africa.

We estimate that application costs are large and, consistently with our initial descriptive evidence, positively correlated with jobseeker ability. For the group of individuals who value the job the most, the correlation between application costs and jobseeker ability is .57. The magnitude of application costs is also substantial. At the mean, application costs amount to 13.5 percent of the monthly wage for the same group of jobseekers. These central findings are robust to the use of different assumptions about the information available to jobseekers, alternative sets of empirical moments, and different restrictions on parameter heterogeneity.[3] Using a local estimate of the value of cognitive ability, we calculate that for the average firm in this market the internal rate of return (IRR) of the application incentive is 9.8 percent. This is much higher than the IRR of the high wage treatment, which is a costlier intervention. However, when we bootstrap the IRR calculation, we find that this policy carries some downside risk. Through counterfactual policy analysis, we show that the IRR increases substantially when the incentive is either (i) targeted to the demographics that drive the treatment effects (e.g. women) or (ii) offered conditional on a good performance on the selection test. Further, these counterfactual policies substantially reduce the downside risk of the original intervention.

Our results make several contributions to the literature. First, we highlight that upfront costs can worsen selection in an important economic context. To our knowledge, ours is the first worker selection experiment that manipulates application costs. Some recent experiments in developing and developed countries have manipulated the wage, or workers' expectations about the wage (Dal Bó et al., 2013; Deserranno, 2019; Belot et al., 2017; Ashraf et al., 2018). These studies typically find that higher wages attract better applicants. For example, Ashraf et al. (2018) show that offering career incentives enables the Zambian government to recruit more talented nurses.[4] A number of papers set in the US have also studied how various contract features affect applications decisions (Flory et al., 2014; Mas and Pallais, 2016). Finally, a recent paper by Hardy and McCasland (2017) studies the experimental placement of apprentices in small firms in Ghana and finds evidence consistent with hiring frictions. None of these studies di-

[3] A potential concern is that unobserved variation in the value of the job may inflate the estimates of application costs and of the cost-ability correlation. While we cannot fully rule out this possibility, we show that our key results are robust to the use of two different estimation strategies that plausibly reduce this unobserved variation.

[4] An exception is Deserranno (2019), who finds that higher expected salaries select less motivated candidates for a non-profit organisation in Uganda.

rectly varies applications costs. Our findings highlight that, when jobseekers find it costly to participate in the labour market, firms may hire better workers if they reduce application costs.

Second, we contribute to a recent, growing literature that studies frictions in the allocation of talent. Previous studies have focused on the role of discrimination (Hsieh et al., 2013), migration costs (Bryan and Morten, 2015; Imbert and Papp, 2016; Lagakos et al., 2017), housing market failures (Hsieh and Moretti, 2015), and corruption (Weaver, 2016). We provide original empirical evidence on the importance of search frictions — in particular, frictions in the job-application process. These frictions have been the focus of several theoretical papers, but direct evidence on their magnitude has been limited to date (Banerjee and Newman, 1993; Marimon and Zilibotti, 1999; Rogerson et al., 2005; Paserman, 2008; Galenianos et al., 2011). A unique feature of our study — and a key contribution to this literature — is that search frictions are identified using exogenous experimental variation.

Our findings are consistent with those of an emerging literature that studies spatial frictions in urban labor markets. This literature shows that across a range of different contexts — including the US, Ethiopia and South Africa — transport subsidies increase job search intensity and impact labor market outcomes (Phillips, 2014; Franklin, 2017; Abebe et al., 2020; Banerjee and Sequeira, 2020). These papers however do not observe the counterfactual workers that would have been attracted by firms in the absence of the program, and thus do not capture changes in the selection of talent.[5] Our study, on the other hand, documents that decreasing application costs enables employers to attract higher-ability applicants, mostly by incentivising a pool of female, inexperienced, unemployed jobseekers that are unlikely to quickly secure good positions otherwise. This suggest that job search assistance policies may have positive impacts on the allocation of talent and motivates the implementation of new market-level evaluations designed to investigate these effects (Crépon et al., 2013).

Lastly, our results highlight that managers do not have accurate beliefs about the returns to different recruitment practices and may thus fail to optimise firms' recruitment policies. Providing information to managers may thus be a cost-effective intervention in this context. These results contribute to the nascent literature on behavioural firms (DellaVigna and Pope, 2016; DellaVigna and Gentzkow, 2017; Kremer et al., 2018).

[5] Crépon and Van den Berg (2016) and McKenzie (2017a) offer recent reviews of the job search assistance literature. Additionally, there is a relevant literature at the intersection of urban and labour economics that studies spatial frictions that result from urban segregation, mostly in developed countries, which is summarised in Gobillon et al. (2007) and Zenou (2009).

2 Context

In this section we describe the labour market in Addis Ababa from the point of view of both firms and workers. We do this by reporting descriptive statistics from our data on jobseekers and firms, as well as from the longitudinal, high-frequency labour market survey collected by Abebe et al. (2020). This evidence suggests that returns to cognitive ability are likely to be substantial in this context, that job search and application costs are an important barrier for many jobseekers, and that high-cost jobseekers may be better selected than low-cost jobseekers.

2.1 The challenge of finding high-ability workers

Finding a worker with the right ability and skills can be challenging for firms in Addis Ababa. To collect data on employers and their beliefs, we sample 196 firms that advertised a vacancy for a clerical job during a period of six weeks in 2017 and ask managers about the HR problems they face and the HR practices they have adopted.[6] The most frequently mentioned HR challenge is finding workers with the right skills. As shown in Figure 1, about 35 percent of managers consider this to be the most pressing HR problem for their firm. Retention, absenteeism, motivation and conduct are all mentioned less frequently than hiring. In terms of HR strategies, about 60 percent of managers report that offering higher wages is the most effective way to improve the quality of recruits. Application incentives are mentioned rarely and, in practice, they are not frequently used by firms in the city.

< **Figure 1 here.** >

The firms in our sample hire workers on a frequent basis. In the two months preceding the interview, the average monthly hiring rate among these firms was 2.5 percent and the average separation rate was 1.6 percent. Hiring thus occurs both to expand the workforce and to replace workers who leave the firm. These labour flows are somewhat

[6]The firms are selected in the following way. First, we screen all vacancies advertised on the main job-vacancy boards or in a popular newspaper insert. To identify clerical jobs, we categorise each vacancy according to the 2010 Standard Classification of Occupations of the US Bureau of Labor Statistics. For the full list of occupations included in the survey see Table A.1. In each firm, we request to interview the head of the selection committee – typically the head of the HR department or the firm's CEO. We use this sample of managers to run the second experiment reported in the paper. During the interview, each manager first completes the CV-ranking and forecast tasks, which we describe in detail in Section 7, and then answers the survey questions about his or her firm.

smaller than those experienced by firms in the US (e.g. in the US, in June 2017, the average monthly hiring rate was 3.7 percent and the average monthly separation rate was 3.6 percent (Bureau of Labor Statistics, 2017)).

Hiring is costly in terms of both time and money. Among firms in our sample, average recruitment costs amount to about 104 USD and 18 hours of staff time (worth about 40 USD when valued at the mean wage of an HR manager in the same firms). Total costs correspond approximately to one month of salary for one of the high-wage jobs in the experiment. These costs do not vary substantially with the number of applicants (many of the costs, such as those related to advertising and developing tests and interviews, are fixed). Managers estimate that considering one more application entails no further monetary costs and would not require more than one hour of staff time.

Firms usually screen workers by assessing CVs and by administering written tests and interviews. Educational qualifications, GPA and previous work experience are the most important variables that managers consider when they assess candidates' CVs. Firms often require applicants to deposit their CV and the other application materials in person. Written tests and interviews are also used frequently. Both interviews and written tests are used to assess general cognitive ability, specific technical knowledge, and personality traits.

2.2 The cost of finding a job and the returns to ability

Finding a job is also challenging in this labour market. First, jobseekers spend substantial amounts of time and money to identify vacancies and apply for them. Using self-reported expenditure data, Abebe et al. (2020) estimate that the monetary cost of searching and applying for jobs amounts to one quarter of weekly expenditure for individuals who are actively looking for employment. To pay for these costs, jobseekers need to frequently take up informal, short-term jobs, which are easier to secure. These challenges are described in detail in Abebe et al. (2020). Here we report one additional piece of descriptive evidence: jobseekers apply for a very small fraction of the available vacancies. In our sample, for example, the average unemployed person in the control group completes approximately two job applications in 30 days. On the other hand, when we screened job boards and newspapers over a similar time period, we were able to find at least 30 relevant vacancies per week. This low number of applications is consistent with the existence of financial constraints that limit job search intensity.

Second, ability matters: highly talented individuals — particularly those with strong cognitive ability — earn more than low-ability individuals. In Table A.2, using data from Abebe et al. (2020), we show that a one-standard-deviation increase in a jobseeker's Raven test score is significantly associated with a 9 percent rise in wages, while similar improvements in conscientiousness or neuroticism are not significantly correlated with wages. Better performance on the Raven test is also significantly related to employment: a one-standard-deviation increase in the Raven score is associated with a 4.5 percentage point (8 percent) gain in the probability of employment. Including standard controls for age and labour market experience does not affect these results (Table A.3).[7]

2.3 Are jobseeker ability and application costs correlated?

Finally, there is evidence that ability and application costs may be correlated among jobseekers.[8] In this last subsection, we provide reduced-form descriptive evidence on this correlation, and on the dynamic selection mechanism that may drive it. Our hypothesis is that, in a given cohort, high-ability individuals who face low application costs find work faster and leave job search earlier than individuals with similar ability who face high application costs.[9] Thus, low-cost, high-ability individuals become progressively under-represented in the pool of jobseekers compared to high-cost, high-ability types. This generates a positive correlation between ability and application costs among jobseekers, which strengthens over time.

We provide evidence on this correlation using the longitudinal data collected by Abebe et al. (2020). This dataset is ideal for this exercise as it provides fortnightly data on the job-search decisions and employment outcomes of a sample of young adults in Addis Ababa for the period of one year. The sample was restricted to individuals who, at the start of data collection, did not have a formal, open-ended work contract. By the end of the year, about half of the people in the sample found employment. The data thus covers a period of active job search and job finding. The data also includes a Raven test administered close to the beginning of the panel and two variables which can proxy for search costs: a measure of financial resources (savings at baseline) and a proxy for

[7] For a systematic discussion on the returns to talent and, more broadly, human capital in developing countries see Porzio (2017) and Caselli et al. (2014).

[8] We use the term 'jobseeker' to indicate those individuals who are actively looking for employment at a given point in time. Thus, in our experimental sample, all individuals are jobseekers, as they have called to inquire about the experiment's position — an active job-search step. In the sample of Abebe et al. (2020), on the other hand, we observe both individuals who look for employment and individuals who do not.

[9] The reverse may happen among low-ability individuals, who have a low chance of finding formal employment. Those who are low ability and high cost may decide to give up on the search for formal work altogether, and restrict themselves to the informal labour market instead. On the other hand, low-ability types that face low costs may invest in formal job search for a longer period of time.

transport costs (distance from the city centre).

We find three pieces of evidence that support our selection hypothesis. First, high-cost, high-ability jobseekers are less likely to find a job than low-cost jobseekers with equal ability. This is in contrast to low-ability individuals, for whom job-finding rates are similar irrespective of application costs. We show this point through the regression analysis reported in Table A.4 in the Appendix, where we study month-to-month transitions from job search to employment. We find that, compared to low-cost types, high-cost individuals experience a lower increase in the probability of finding work for the same increase in ability — a significant effect when costs are proxied by savings and an insignificant effect of a roughly similar magnitude when costs are proxied by distance. Among jobseekers with ability one standard deviation above the mean, high-cost types have a probability of finding a job that is 8 percentage points lower than low-cost types. On the other hand, among jobseekers with ability one standard deviation below the mean, high and low-cost types have very similar job-finding rates (about one percentage point apart).[10]

Second, the correlation between jobseeker ability and application costs strengthens over the course of the year. In Figure 2 and Figure 3, we plot the fortnightly value of the cost-ability correlation in the selected sample of jobseekers. When costs are proxied by savings, we estimate that the correlation between jobseeker ability and application costs grows by a significant 0.011 of a standard deviation every fortnight. When costs are proxied by distance from the city centre, the correlation grows by a significant 0.018 of a standard deviation every fortnight. In other words, there is a growing gap in average ability between low and high-cost jobseekers (which we also illustrate showing the separate trends of high and low-cost types in Figure A.1 in the Appendix).

< Figures 2 and 3. >

Third, by the end of the year, the correlation between jobseeker ability and application costs is large and positive. As we show in Figure 2, this correlation is large (and just marginally insignificant) for our first proxy of applications costs — low savings. Among the individuals who search for work in the last fortnight of the panel, those with below-median savings have a Raven test score that is .3 standard deviations higher than those

[10] These figures refer to the estimates reported in column 1 of Table A.4, where we proxy application costs by savings. Individuals with high savings may also have higher reservation wages. In the data, a one standard deviation increase in savings is associated with self-reported reservation wages that are about 0.15 standard deviations higher. Higher reservation wages should reduce the job finding rates of high saving individuals, partly cancelling the effect of their lower search costs. Thus, the results on job finding that we report here likely underestimate the differential selection that is due to search costs alone.

with savings above the median. In the last fortnight of the panel, the correlation with our second proxy — distance from the city centre — is also positive, but it is smaller and less precisely estimated than the correlation with savings (Figure 3). Jobseekers who live above the median distance from the city centre have a Raven test score that is an insignificant .08 standard deviations higher than jobseekers who live below the median distance.

Overall, this evidence supports the hypothesis that selection dynamics generate a positive correlation between jobseeker ability and application costs. This observation motivates a model that explicitly considers how such a correlation can affect the hiring outcomes of a firm, which we present in the next section.

3 A simple model of job application decisions

We propose a simple model of application decisions that captures two key frictions in job search: application costs and uncertainty about the probability of being offered the job. The model describes the effects of application incentives on application rates and on the quality of the applicant pool.

Jobseeker Characteristics. Consider a set of individuals deciding whether to apply for the experiment's job. For tractability, let us focus on the large-number case and assume that these jobseekers form a continuum of unit measure.

Jobseekers differ in terms of their ability (noted T in what follows); as well as in terms of the benefit that they derive from being offered the job (noted B). Heterogeneity in B captures differences in outside options. To fix ideas, it is helpful to think of T as the score on the Raven test (a reliable predictor of worker performance) and of B as the monetary net present value of being offered the job (where a negative net present value translates into $B = 0$, since being offered the job does not require jobseekers to *take* the job). These are the empirical counterparts that we use for estimation, as described in Section 6.

Jobseekers who wish to apply must incur a cost (noted C), which is heterogeneous across the population. C is the net opportunity cost of applying for the job, that is, the economic value of all the things that jobseekers have to give up in order to apply—typically both money and time.[11] This cost is heterogeneous for two reasons. First, the time and money required to make the application differ across jobseekers (e.g. jobseek-

[11] We also allow C to be negative. This captures the fact that some people may derive a net benefit from attending the testing sessions, independently of getting the job (e.g. because of the value of networking, or because they learn something valuable about the market).

ers who live farther away from the application centre have to pay a more expensive bus fare). Second, the value of time and money differs according to the circumstances of the jobseeker (e.g. poorer jobseekers will find it relatively more expensive to pay the same bus fare compared to jobseekers with better financial resources). If jobseekers dislike being tested, C will also include the psychic or hassle costs of the application.

We characterise jobseekers along these three dimensions: (T, C, B). For simplicity, we assume a finite number of 'benefit types', so that ability T and application costs C have a continuous joint distribution for each of these B-types. More precisely, we make the following assumptions about the distribution of characteristics across the population of jobseekers.

Assumption 1 The benefit from receiving a job offer is given by

$$B \in \{b_1, b_2, ..., b_n\} \text{ where } b_z \geq 0 \text{ for } \{z = 1, 2, ..., n\}.$$

Assumption 2 Conditional on $B = b_z$, ability T and application costs C follow a bivariate normal distribution

$$\begin{pmatrix} T_z \\ C_z \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_{T_z} \\ \mu_{C_z} \end{pmatrix}, \begin{pmatrix} \sigma^2_{T_z} & \rho_z \sigma_{C_z} \sigma_{T_z} \\ \rho_z \sigma_{C_z} \sigma_{T_z} & \sigma^2_{C_z} \end{pmatrix}\right] \; for \; \{z = 1, 2, ..., n\}.$$

Throughout the rest of the paper we use the same notation introduced in Assumption 2. That is, we use sub-indices to denote ability and costs conditional on B-types. Further, we use σ_{CT_z} to denote the covariance between C and T, and ρ_z to denote the correlation between these two variables.

Application incentive and wage subsidy. We model the application incentive as a shock that lowers application costs, shifting the distribution of C to the left by an amount τ. Similarly, we model the wage subsidy as a shock that raises the value of the job, shifting the distribution of B to the right by an amount τ^w. In both cases the assumption is that, in line with our empirical findings, the interventions reduce the cost-benefit ratio without affecting jobseekers' beliefs about the probability of being offered the job upon applying.[12]

[12]We present empirical evidence showing that beliefs about the probability of being offered the job are not affected by treatment in Section 5.5 and Table A.40. The fact that beliefs do not respond to treatment is consistent with the finding that, in a beauty contest game played during the application process, 80 percent of applicants are not strategically sophisticated (Crawford et al., 2013).

Selectivity and information. When jobseekers apply for the job, the employer observes their ability T. Jobseekers, accordingly, make application choices on the understanding that they will get the job if $T > a$, where a captures the perceived selectivity of the recruitment process.

Jobseekers face uncertainty about the likelihood of being offered the job conditional on application. We assume jobseekers know the cost of applying, as well as the benefit of getting the job; but they face uncertainty about recruitment outcomes because they have either imperfect information about the recruitment policy, or imperfect information about their ability. We refer to these two information benchmarks as the 'noisy selection' and 'noisy ability' cases.

In what follows, we show that in both cases the correlation between cost and ability plays a key role in determining the impacts of application incentives. For a sufficiently large positive correlation, application incentives raise applicant ability regardless of whether the source of uncertainty is ability or selectivity. However, these similar predictions are underpinned by selection decisions made on the basis of different information. In the empirical analysis, we will thus estimate the model under both benchmarks and probe the robustness of our quantitative findings to the these different assumptions.

3.1 Analysis

The noisy-selection case. To model this case, we assume that ability is known, but selectivity is observed with noise: jobseekers anticipate that the threshold necessary to get the job is a normally distributed random variable with mean μ_a and variance σ_a. Thus, an individual with ability t believes that the probability of being offered the job conditional on an application is $\Phi\left(\frac{t-\mu_a}{\sigma_a}\right)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

B-type jobseekers with cost $C_z = c_z$ and ability $T_z = t_z$ will apply for the job if and only if

$$\Phi\left(\frac{t_z - \mu_a}{\sigma_a}\right) \geq \frac{c_z}{b_z} \tag{1}$$

From equation 1 it is easy to see that if T_z and C_z are not correlated, then an application subsidy that shifts the cost-benefit ratio without affecting beliefs about selectivity, will lead to more but worse applications on average. This is the standard logic that often guides selection policies—a costly application process can improve selection by discouraging the participation of unwanted candidates, who are privately informed about their low ability. Thus, subsidising application costs may reduce average applicant ability.

Proposition 1. *Suppose (T, B, C) are observable and distributed according to Assumptions 1 and 2, with $\rho_z = 0$. Further assume jobseekers anticipate that the threshold necessary to get the job is $a \sim \mathcal{N}(\mu_a, \sigma_a)$. Then it follows that for each $B = b_z > 0$, the application incentive (i) increases application rates, and (ii) decreases the average ability of applicants.*

Proof. Within each B-type, jobseekers who receive the application incentive face a positive shock that lowers their application costs by an amount τ. This means that the application rate among these jobseekers is given by

$$\Pr\left(C_z \leq \Phi\left(\frac{T_z - \mu_a}{\sigma_a}\right) b_z + \tau\right) \tag{2}$$

Similarly, the expected ability of applicants in this group is given by

$$E\left(T_z \mid T_z \geq \Phi^{-1}\left(\frac{C_z - \tau}{b_z}\right) \sigma_a + \mu_a\right) \tag{3}$$

If C_z is not correlated with T_z, then it trivial to check that the first expression is increasing in τ, and the second expression is decreasing in τ. $\square$

From equations 2 and 3 it is easy to see that the application incentives generally operate through two related but different channels. On the one hand, the application incentives weakly reduce the ability of marginal applicants at any given level of costs; and on the other hand, the incentives attract additional high-cost applicants at any given level of ability.

When ability and costs are not correlated (i.e. when $\rho_z = 0$) only the first effect matters to determine the impact of the intervention on average applicant ability—this is the standard intuition, captured in Proposition 1. Reducing the cost-benefit ratio increases the number of applications, but lowers the average quality of the applicant pool.

In contrast, when $\rho_z > 0$, application incentives activate two forces that act in opposite directions. Lowering the cost-benefit ratio increases applications by attracting a larger share of high-cost applicants, but these new applicants tend to have higher ability on average. If the correlation between ability and costs is strong enough, the second channel dominates and the intervention has a positive selection effect.

In Figure 4 we illustrate this intuition graphically. The application region is shaded in blue $\left(\frac{C_z - \tau}{b_z} \geq \Phi\left(\frac{T_z - \mu_a}{\sigma_a}\right)\right)$, while the superimposed purple contours show the density of the joint distribution of jobseekers $(\Pr(T_z = t_z, C_z = c_z))$. As the figure shows, the application incentive shifts the application threshold to the right, for all levels of ability. However, given the positive correlation between T_z and C_z, this expansion of applications adds a larger share of high-cost, high-ability jobseekers. As a consequence, the average ability of applicants increases.

Figure 4 also highlights that, when cost and ability are not correlated, we would expect applicants to have higher average ability than non-applicants. However, the opposite may be true for a sufficiently strong cost-ability correlation. Indeed, in our experimental data, non-applicants have on average higher ability than applicants (as measured by their GPA score). We will come back to this point in the structural section and show that, while we do not target this moment explicitly, the model nonetheless correctly reproduces it.

< Figure 4 here. >

Proposition 2. *Suppose (T, B, C) are observable and distributed according to Assumptions 1 and 2. Further assume jobseekers anticipate that the threshold necessary to get the job is $a \sim \mathcal{N}(\mu_a, \sigma_a)$. Then it follows that for each $B = b_z > 0$, the application incentive (i) increases application rates, and (ii) increases the average ability of applicants, whenever*

$$\frac{\sigma_{T_z}}{\sigma_a \sigma_{C_z}} \frac{b_z}{\sqrt{2\pi}} \leq \rho_z$$

Proof. See Appendix A.1

Proposition 2 captures the intuition behind our main experimental result. The application incentive attracts a group of marginal applicants who face larger application costs compared to control group applicants; and because the correlation between costs and ability is positive among jobseekers, these marginal applicants have, on average, higher ability than the applicants in the control group. If the cost-ability correlation is sufficiently large, this indirect channel dominates the standard channel presented in Proposition 1, and the application incentive raises average applicant ability.

This proposition also illustrates the key role played by uncertainty. As noise σ_a grows, the condition on the cost-ability correlation becomes weaker. Intuitively, when the outcome of an application is highly uncertain, application costs play only a limited screening role. In this case, the 'standard intuition' effect is small. Thus, a moderate indirect effect is sufficient to obtain a positive impact on applicant ability.

The noisy-ability case. Let us now consider the second scenario. Here we assume jobseekers are confident that the selection threshold is fixed at some level a, but they do not directly observe their ability.[13] Since C and B are known, and are potentially informative about ability, we assume that jobseekers can use these variables to update their belief about the probability of being offered the experiment's job. To fix ideas on this set of assumptions, consider that jobseekers can observe the application conversion rates of people with similar observables, and can use these to make inferences on the probability of a job offer. For example, consider an individual who lives in a central neighbourhood (low C) and has a strong educational record (high expected wage/low B). Under our assumptions, this person believes that their chances of being offered the experiment's job are the same as those of other applicants from her neighbourhood with a similar educational record. Formally, we have that B-type jobseekers with cost $C_z = c_z$ will apply if and only if $c_z \leq c_z^*$, where c_z^* is the level of costs for which

$$\Pr(T_z > a \mid C_z = c_z^*\,,\, B = b_z) = \frac{c_z^*}{b_z} \tag{4}$$

From equation 4 it is clear that in this case the 'standard intuition' that we discussed before is muted. Jobseekers do not directly observe T_z, so they are not able to self-select directly on ability at the application stage. In fact, if C_z is not correlated with T_z, then subsidising applicants would increase the number of applicants, but these new applicants would be a random selection from the pool of jobseekers; so the application incentive would yield no impact on the expected ability of applicants.

Importantly, however, when $\rho_z > 0$ the indirect channel still operates, because the incentive still attracts on the margin applicants who face larger application costs compared to control group applicants. So in this case too, if costs and ability are positively correlated, the application incentive intervention can yield a positive selection effect. In Figure 5 we illustrate this graphically.

< Figure 5 here. >

The application region is again shaded in blue ($c_z < c_z^*$), and the superimposed purple contours show the density of the joint distribution of jobseekers ($\Pr(T_z = t_z, C_z = c_z)$). As the figure shows, the application incentive shifts the application threshold to the right, for all levels of ability. However, given the positive correlation between T_z and

[13] Jobseekers can be wrong about the selection threshold due to overconfidence as in Spinnewijn (2015) and Banerjee and Sequeira (2020). Further, as explained above, we assume that jobseekers do not revise their beliefs about a in response to treatment. It is possible to show that if we assume instead that a is the true selectivity threshold, which responds endogenously to treatment, application incentives increase applicant quality if and only if the cost-ability correlation is positive. In Proposition 3 below, we show that a positive cost-ability correlation is similarly required for the case where a is exogenous. So our key results here are not driven by this assumption.

C_z, this expansion of applications adds a larger share of high-cost, high-ability jobseekers; and as a consequence, the average ability of applicants increases.

Proposition 3. *Suppose (T, B, C) are distributed according to Assumptions 1 and 2. Assume jobseekers are confident about the selection threshold a, but they only observe C and B, which they can use to update their beliefs about the probability that they will pass the recruitment test $T > a$. Then for each $B = b_z > 0$, the application incentive (i) increases application rates, and (ii) increases the average ability of applicants, if and only if*

$$0 < \rho_z < \frac{\sqrt{2\pi}\sqrt{1-\rho_z^2}\sigma_{C_z}}{b_z}$$

Proof. See Appendix A.1

Taken together, Propositions 2 and 3 show that the pivotal role of the cost-ability correlation does not depend on whether the uncertainty about the outcome of the application stems from imperfect information over ability, or imperfect information over the recruitment policy. On the other hand, the nature of uncertainty can affect the selection decision in other ways. For example, the 'standard intuition' effect only exists in the case where jobseekers know their own ability. Empirically, we do not know which set of assumptions better approximates the information available to jobseekers. Thus, in the structural estimation section, we will bring both information benchmarks to the data and probe the robustness of our findings.

4 Design and data

4.1 Design

We study the recruitment of workers for clerical jobs in Addis Ababa. The experiment takes place over eight consecutive fortnights. On the Sunday at the beginning of each fortnight, the positions are advertised in a local newspaper and in the main job vacancy boards of the city. The advertisement describes the position as a three-months fixed term appointment based in Addis Ababa and specifies that candidates must hold a university degree or a vocational diploma. Interested individuals are invited to call a specified phone number to get more information about the position and the application process. The deadline for applications is on the Friday of the same week. The positions are based at a local organisation specialised in research and data collection.[14]

[14] At the time of the experiment, the organisation employed about 60 permanent workers and 50 fixed-term-contract workers with a similar profile as those recruited for the experiment. All workers hired

A small team of enumerators answers the phone calls of interested jobseekers following a standardised script. First, they ask a short number of questions capturing callers' socio-demographic characteristics and work experience. Second, they give some information about the position. Third, they explain that, in order to apply for the position the jobseeker has to attend a testing session at our application centre, on a specified day. Jobseekers have to bring to the session a CV, a cover letter and proof of identity.

We randomly vary two attributes of the position across callers: the wage and whether we offer an application incentive. Callers assigned to the *control group* are informed that the position pays a monthly wage of 1,600 ETB (74 USD) and are not offered the application incentive. Callers assigned to the *application incentive* group are also told that the position pays a wage of 1,600 ETB per month. In addition, these callers are informed that, if they complete the testing session, they will receive a monetary payment of 100 ETB (4.5 USD). This payment is presented as a reimbursement of the costs jobseekers may incur in the application process. Finally, callers assigned to the *high wage* group are told that the position pays a wage of 3,200 ETB (148 USD) per month and are not offered the application incentive. We calibrated these wages at the 35th and 75th percentile of the distribution of earnings for similar positions using data from Abebe et al. (2020). Using jobseekers' assessment of the probability of getting the job, we calculate that the expected value of the high wage offer is worth about 105 USD for the average subject.

All jobseekers who call before the application deadline on a given fortnight are assigned a testing day.[15] This can be from Monday to Friday of the second week of that fortnight, or on the first Monday of the following fortnight. To reduce the risk of contamination across experimental conditions, individuals assigned to different treatment groups are invited to take the test on different days.[16] Each treatment group is assigned

as part of the experiment were employed for three months and were paid the same wage (which corresponds to the wage offered in the high wage condition discussed below). This high wage was a surprise for those workers selected under a low wage offer. The fact that the actual wage would differ from the initial offer was not disclosed to any of the staff involved in the implementation of the surveys and the other experimental activities. The same organisation was contracted for the data collection required for the other parts of this study (e.g. the survey of other employers to elicit beliefs).

[15]We do not allow jobseekers to call on more than one fortnight. After each phone call, enumerators check our database and disqualify the person if they have called on a previous fortnight.

[16]To further reduce the risk of contamination, callers that are told tha the employer is hiring for multiple positions. If callers assigned to different treatment groups talk to each other about position, this feature should help them explain why different callers are offered different terms. Specifically, callers in the control group are told that they have been assigned to a position called 'position A'. Callers in the application incentive and high wage groups are informed that they have been assigned to positions 'B' and 'C', respectively. We do not give any information about why a jobseeker is assigned to a particular position. If asked, the enumerator will reply that (i) the enumerator is not authorised to disclose the ex-

to two of these six testing days. This assignment of treatments to testing days is randomly varied every fortnight. If a jobseeker cannot attend the testing session on the proposed day, we allow them to attend the other testing session assigned to his or her treatment group for that fortnight.

We call back all jobseekers four weeks after the first phone call. In this second interview, we ask a set of questions about the job applications that individuals have made in the 30 days after the first phone call. Completion of this second phone interview is incentivised with a monetary payment of 20 ETB (.85 USD). We show the full timeline of a typical hiring round in Figure A.2 in the Appendix.

We offer three jobs per fortnight – one per treatment group.[17] For each position, the five applicants with the highest score on an index of cognitive ability (which combines the scores on the Raven and Stroop tests) are invited for an interview. The employer decides who among these interviewees is given the job.

As much as possible, the experiment follows the usual recruitment protocols of the employer, which are similar to those of other employers in this market. Importantly, in Addis Ababa it is standard practice not to include information about the wage in the initial advertisement for a position. Further, as explained in Section 2, written tests like the one we employ are common.

4.2 Measuring applicant ability

We measure the ability of the individuals that apply for the experiment's job with a number of popular personnel selection tests. These tests are good predictors of worker productivity and are routinely used by firms worldwide (Chamorro-Premuzic and Furnham, 2010). We also collect information about relevant work experience and GPA scores – two variables that local employers use to screen applicants. Finally, we also collect information about economic preferences.[18]

We measure cognitive ability with the widely used Raven and Stroop tests. The Raven test measures fluid intelligence, the ability to make meaning out of complex information and to reproduce this information. Several meta-analyses have identified the Raven test as the single best predictor of worker productivity (Schmidt and Hunter, 1998; Raven, 2000). This test has been widely used in the recent economics literature to measure worker ability (Dal Bó et al., 2013; Beaman et al., 2013; Abebe et al., 2020). The

act criteria we use to assign callers to positions, (ii) one major factor is to keep the number of applicants across positions constant.

[17]In a small number of instances, we combine two fortnights of the same treatment group together. In this case, we offer only one job to the applicants assigned to that treatment group in these two fortnights.

[18]We provide a detailed discussion of each measure in Appendix A.3.

Stroop test is a popular test of cognitive control, the ability to direct and discipline attention which is required to perform complex tasks (Diamond, 2013). We use a version of the Stroop task developed by Mani et al. (2013).

For non-cognitive skills we use two widely used and validated scales: the big five inventory (BFI-44) and the grit scale (John and Srivastava, 1999; Duckworth et al., 2007). We focus on three facets on non-cognitive ability which have been identified as particularly relevant to work performance: conscientiousness, neuroticism and grit. These respectively capture a careful and vigilant attitude at work, the ability to deal with stressful situations, and the capacity to persevere through challenges (Chamorro-Premuzic and Furnham, 2010). We perform standard validity checks for the psychometric measures and satisfy accepted thresholds (e.g. see Table A.5 for Cronbach α). Laajaj and Macours (2017) emphasise the value of performing validity tests when psychometric scales are used in new contexts. We also administer scales measuring locus of control and confidence.

Further, we collect information about relevant work experience. For this purpose, we use the classification of tasks developed by Autor and Handel (2013). This includes the following categories: physical, routine, problem-solving, managerial, mathematical, and client-interaction tasks. For each of these, we ask participants to report the number of months of experience in jobs that required them to perform that task on a regular basis. We focus on routine, problem-solving and managerial tasks, as these were identified by the employer as the most relevant types of experience.

We aggregate the individual measures in indices of cognitive ability, non-cognitive ability and experience. Each index is constructed as the sum of the standardised values of three measures, as reported in Table A.6 in the Appendix (Anderson, 2008).[19]

Finally, we measure four types of economic preferences: an incentivised measure of time preferences, and non-incentivised measures of risk preferences, social preferences and level-k rationality. The task to measure time preferences is an adapted version of the game by Augenblick et al. (2015). In this task, participants have to allocate units of work across two work sessions. For risk preferences and social preferences we use questions from the Global Preferences Survey (Falk et al., 2016). Finally, we administer a simplified and non-incentivised version of the beauty contest game to elicit level-k

[19]We think of the three components of the index as representing three distinct facets of a particular ability. We thus give each component of the index equal weight. Results, however, are qualitatively unchanged if we weight by the inverse of the covariance matrix. Further, ideally, we would standardise each test score by the variance of that test for a pool of applicants attracted at the average wage. We are unable to do this, as we either attract applicants at a below-average or at an above-average wage. We thus use the variance of a given test among all applicants in order to standardise the value of that test.

rationality (Crawford et al., 2013).

4.3 The sample, randomisation and attrition

Over the eight fortnights of the experiment, 4,689 jobseekers called to inquire about the position — an average of about 590 individuals every fortnight. The number of fortnightly callers stayed constant over the course of the experiment, suggesting that that the positions generated sustained interest among jobseekers. Table 1 reports summary statistics for the population of individuals who called to inquire about the position. The typical caller is young, male and has some work experience. The average age is 26. 15 percent of the sample is 30 or older. Women account for 21 percent of the sample. On average, callers have 28 months of wage-work experience. This masks substantial heterogeneity, as 47 percent of the sample has no work experience. Callers also have a variety of educational backgrounds.

< Table 1 here. >

We randomise the offer of the two treatments using a stratification rule in order to improve covariate balance (Bruhn and McKenzie, 2009). We create strata of six consecutive callers of the same gender and same level of work experience.[20] In each stratum, we randomly allocate two callers to the control group, two callers to the application incentive group and two callers to the high wage group. These callers are invited to a testing session at our application centre during the following week. There are two testing sessions per treatment group, per fortnight. We randomise the allocation of testing sessions to days of the week. We do this in a single draw for all eight fortnights and re-randomise until we have an allocation that is balanced across days of the week.[21]

We find that covariates are balanced across treatment groups and that attrition is modest and uncorrelated with treatment. 1,557 callers are assigned to the control condition, 1,559 to the incentive condition, and 1,573 to the high wage condition. Table 1 reports descriptive statistics and balance tests for the characteristics of the callers that we measure during the first phone interview. The joint orthogonality tests show that, overall, characteristics are balanced across treatment groups. Individual covariates are

[20]We define an experience dummy using the median number of months of work experience of callers in the pilot.

[21] The randomisation rule is that (i) each treatment should be allocated two testing days each fortnight, and (ii) no treatment should be allocated, overall, more than three or less than two sessions on the same day of the week. For this exercise, we consider the Monday session on the the following fortnight as being a distinct 'day of the week'.

also generally balanced.[22] Further, appendix Table A.9 shows that the assignment of individuals to treatment is strongly balanced across weeks. In terms of attrition, in the second phone survey, we interview 93.5 percent of the sample (attrition is thus 6.5 percent). This is consistent with recent studies with similar populations in urban East Africa (Abebe et al., 2020). Figure A.3 shows that attrition is not systematically related to treatment status and Table A.8 confirms that the sample remains balanced after attrition.

Using data from the 2013 Labour Force Survey (LFS), we also show that our experimental sample resembles in many ways the population of jobseekers in Addis Ababa. In particular, average age and work experience in our sample match those of a comparable sample of jobseekers from the LFS, as we show in Table A.10 in the online Appendix. Further, in Figure A.4, we show that the distribution of age in the two samples is qualitatively similar and statistically indistinguishable. We find some differences in terms of gender (women are under-represented in our sample) and unemployment duration (the long-term unemployed are under-represented in our sample). These differences are likely due to the fact that women and the long-term unemployed search at lower intensity and hence are less likely to respond to the ad placed by the employer. As we will show in Section 5, these groups of jobseekers tend to drive the treatment effects on applicant quality. Our estimates are thus likely to be a lower bound of the treatment effects that would be observed in a more representative sample.

4.4 Empirical strategy

Our objective is to study the impacts of the interventions on application rates and on the ability of applicants.[23] We estimate effects on application rates using a regression model of the following form:

$$\texttt{apply}_i = \beta_0 + \beta_1 \cdot \texttt{incentive}_i + \beta_2 \cdot \texttt{high wage}_i + \kappa_b + u_i, \tag{5}$$

where $\texttt{apply}_i$ is a dummy that captures whether person i has applied for the job, $\texttt{incentive}_i$ and $\texttt{high wage}_i$ identify individuals who have been offered the application incentive and the high-wage treatment, and κ_b are stratum dummies (Bruhn and McKenzie, 2009). The coefficients β_1 and β_2 capture the change in application rates generated by the application incentive and the high wage offer. This model is estimated over the *baseline*

[22]The only exception is a 3 percentage point difference between the high wage group and the control group in the proportions of callers who are unemployed and wage employed.

[23]All reduced-form analysis was pre-registered. We list in the Appendix a number of variations from our plan motivated by data and fieldwork challenges.

sample, that is, over the sample of individuals who called to inquire about the position. We use a similar model to study the effects of the interventions on expectations and other job-search activities. These outcomes were measured during the second phone call and thus we use the second phone call sample to run this additional analysis.[24]

We study impacts on the ability of applicants by measuring changes in the conditional mean and conditional quantiles of the measures of quality discussed in the previous section. For this analysis, we restrict the sample to all applicants. We measure changes in mean applicant ability by estimating a simple OLS regression model of this form:

$$\mathtt{y}_i = \delta_0 + \delta_1 \cdot \mathtt{incentive}_i + \delta_2 \cdot \mathtt{high\ wage}_i + u_i, \tag{6}$$

where y is a measure of worker ability. Further, we estimate a conditional quantile function of the following form, using a standard quantile regression model (Koenker and Hallock, 2001):[25]

$$Q_\theta(\mathtt{y}_i|X_i) = \gamma_0 + \gamma_1 \cdot \mathtt{incentive}_i + \gamma_2 \cdot \mathtt{high\ wage}_i. \tag{7}$$

For each measure of worker quality y, γ_1 and γ_2 capture the change in conditional quantile θ caused by the treatments. For example, suppose that we are studying the 90th percentile of the distribution of cognitive ability and that we obtain an estimate of γ_1 of 1. This would say that an applicant at the 90th percentile of the distribution in the incentive group has a cognitive ability score that is one point higher than an applicant at the 90th percentile of the control distribution. A key implication of this quantile shift is that the proportion of applicants who score above the 90th percentile of the control distribution increases. This suggests that to study changes in applicant quality we can also compare the probability that an applicant scores above a given threshold across the two groups. In the results section, we show that our findings are robust to the use of this alternative empirical strategy.

We focus the quantile analysis on five percentiles: 90th, 75th, 50th, 25th, 10th. We also present a test of stochastic dominance first proposed by Barrett and Donald (2003). Stochastic dominance occurs when the CDF of one distribution is weakly smaller than the CDF of the other distribution at all points of the support *and* strictly smaller at least at one point. The null hypothesis of the Barrett and Donald (2003) test is that the CDF

[24] As discussed in the previous section, attrition between the two surveys is limited and uncorrelated with treatment. Furthermore, the sample remains balanced after attrition.

[25] In models (6) and (7) we do not include the randomization block dummies. This is because we estimate these models on the sample of applicants, which, by design, does not include all of the individuals that were originally randomised into the three experimental conditions.

of one distribution is weakly smaller than the CDF of the other distribution. To have evidence that distribution A dominates distribution B, we should thus both (i) reject that B is weakly smaller than A and (ii) fail to reject that A is weakly smaller than B. In the results section, we report and interpret the findings of both tests.

We report robust standard errors throughout the paper.[26] Further, we control for multiple-hypothesis testing by using sharpened q-values that control the false discovery rate (Benjamini et al., 2006).[27]

5 Results

5.1 Impacts on application rates

We find that the incentive has a large and significant effect on applications. Individuals in the incentive group are 11.5 percentage points more likely to apply for the position than individuals in the control group. 41 percent of subjects in the control group apply for the position, so this treatment effect amounts to a 27 percent increase in application rates. Further, we find that individuals in the high-wage group are 18.7 percentage points more likely to apply to the position. Thus the application incentive generates an increase in applications that is about two thirds of the increase in applications that can be obtained by doubling the wage. The two effects are statistically different from each other. We report these results in Table 2.

< Table 2 here. >

As shown in Figure A.5, impacts on application rates are stronger in the upper part of the GPA distribution. For example, the increase in applications generated by the incentive treatment in the lowest decile of the GPA distribution is close to 8 percentage points. In the top decile of the distribution, the increase is a significant 14.5 percentage points. These observations suggest that the overall effect on the quality of the pool of applicants, both at the mean and at the top of the distribution, is positive. We discuss these results in detail in the next subsection.

[26]For quantile regressions, robust standard errors are computed using the methods proposed by Machado and Silva (2000) and the Stata command developed by Machado et al. (2011).

[27]We have three indices of ability, so each individual test is repeated three times, forming a family of hypotheses. To calculate q-values, we first compute standard p-values for each test in a given family. Then, we run the sharpened procedure proposed by Benjamini et al. (2006) using these p-values. The q-values we obtain express the expected proportion of false discoveries that we need tolerate if we want to reject the null hypothesis of a given test.

5.2 Impacts on the quality of the applicant pool

The application incentive improves the quality of the applicant pool. This is our most important finding. The incentive raises average cognitive ability among applicants by .25 points, or .12 of a standard deviation (Table 3). This effect is significant at the 5 percent level and is robust to the correction for multiple comparisons.[28] Applicants in the incentive group perform significantly better in both the Raven and the Stroop tests. Compared to applicants in the control group, they answer correctly 1.2 additional questions in the Raven test and they require 2.6 fewer seconds to complete the Stroop task. These treatment effects are of a broadly similar magnitude as those documented in previous worker selection experiments. For example, Dal Bó et al. (2013) document an increase in performance on the Raven test of about half a correct answer. We report the full results for the individual tests in Table A.11 in the Appendix. We also find that the applicants attracted by the incentive have GPA scores that are a significant .1 standard deviation higher than control applicants (Table A.12). This is an important result as many firms in Addis Ababa use GPA scores to assess candidates' ability. Thus the applicant pool improves also in terms of the screening criteria used by local firms.

The increase in quality occurs both at the top and at the bottom of the distribution. These results hold when looking at quantile shifts, number of top and bottom applicants, and average quality of the top applicants. We look at each of these in turn. First, we find that the cognitive ability scores at the 90th, 75th and 25th percentiles improve significantly (Table 3). These effects correspond to about .1 a standard deviation of the cognitive ability index; q-values are always below .15. We also estimate positive, but insignificant effects at the 50th and 10th percentiles. Overall, a Wilcoxon rank-sum test rejects the equality of the cognitive ability distribution in the control and incentive groups (p=.038). Second, we document a large effect on the number of top applicants (defined as individuals above the 90th percentile of the cognitive ability score in the control group). Top applicants nearly double from 63 in the control group to 117 in the incentive group. This effect is generated by a combination of higher application rates, and a significant, 5.3 percentage points increase in the proportion of top applicants in the applicant pool (see Table A.13 in the Appendix). At the same time, the number of applicants at the bottom of the distribution is fairly stable. For example, compared to the control condition, the application incentive attracts only nine additional applicants who score below the 10th percentile of the control distribution. Third, we find that the

[28]This estimate is also not sensitive to the exclusion of tests carried out on specific days of the week. We show this in Figure A.7 in the Appendix. Also, the findings in this section are qualitatively unchanged when we weight the index by the inverse of the covariance matrix (Table A.17,A.18, A.19).

average ability of top applicants increases significantly. In Table A.14 we show regressions of average cognitive ability and GPA scores for samples comprising the top 20, 10 and 5 applicants for each job (the top 5 applicants are invited for the interview). We document sizeable increases in both scores between .2 and .5 standard deviations. In particular, the application incentive increases the GPA scores of the top 5 applicants by a significant .5 standard deviations and the cognitive ability score by a large but insignificant .3 standard deviations.

Consistently with the results for specific quantiles, we find suggestive evidence that the cognitive ability distribution among treated applicants stochastically dominates the control distribution. This is an attractive feature if the firm's objective is to maximise the ability of its hires.[29] We see the characteristic pattern of stochastic dominance when we plot the cumulative distributions of cognitive ability for the two groups (Figure 6). We find a similar pattern if we look at the results of the Raven test, the Stroop test and GPA separately (Figures A.9, A.10 and A.11 in the Online Appendix). Using the formal test of Barrett and Donald (2003) we find no evidence to reject the hypothesis that the CDF of the incentive distribution is weakly smaller than the CDF of the control distribution (p=.949). This result is consistent with dominance of the incentive distribution over the control distribution. However, it also consistent with the equality of the two distributions. We thus also test the null hypothesis that the CDF of the control distribution is weakly lower than the CDF of the incentive group. For this test we obtain a p-value of .136, giving us suggestive evidence of stochastic dominance.

The high-wage offer also attracts an applicant pool with higher cognitive ability. We estimate significant positive effects at the mean, and at the 90th, 75th and 25th percentiles (see Appendix Figure A.12 for the full distribution). The magnitude of these point estimates is smaller than those we obtained for the application incentive, but we cannot reject the null hypothesis that the two treatments have the same effect. The significant estimates of the impact of the high-wage offer are associated with q-values above .1 (and in two cases above .2), suggesting these results are generally not robust

[29] Stochastic dominance makes it possible to unambiguously rank distributions for objective functions that are increasing in the value of the random variable (Deaton, 1997; Barrett and Donald, 2003). Thus, in our setting, the dominant distribution would be preferred both by firms who maximise the expected quality of hires, and by 'risk-averse' firms with an objective function that is increasing and concave in quality. The comparison would not be unambiguous, however, if firms value having a smaller pool of applicants or if acceptance rates are lower in the dominant group. We consider the first point in Section 6. Regarding the second point, we show below that the increase in quality generated by the incentive is concentrated among those jobseekers with the weakest outside options. These jobseekers are likely to have the highest acceptance rates. This further increases the value of the applicant pool attracted by the application incentive.

to the multiple-comparisons correction.

< Table 3 here. >

< Figure 6 here. >

Lastly, we are unable to find significant differences in non-cognitive ability or experience between applicants in the incentive group and applicants in the control group. The high-wage offer significantly increases median non-cognitive ability, but does not significantly affect the other percentiles of the distribution. Tables A.15 and A.16 report the results from these regressions.[30]

5.3 Impacts on search for other jobs and job-search outcomes

We do not find evidence that the application incentive distorts individuals' search for other jobs or impacts their labour market outcomes. This is not surprising, as the small cash incentive ensures that applying to the experiment's job does not deplete the resources available to search for other positions. To study the search for other jobs, we use the data collected during the second phone interview, 30 days after the initial phone call, and a regression model with same form as model (5). We investigate whether the interventions change the number of applications made, the amount of money and time spent on job search, the number of interviews and job offers obtained, and whether the jobseeker is currently working in a new job. We report these results in Table A.21 in the Appendix.[31] For the application incentive, we consistently estimate small and insignificant coefficients.

On the other hand, we find that individuals in the high-wage group have significantly worse outcomes than the controls: they obtain .04 fewer interviews, .03 fewer offers and are about 2 percentage points less likely to be working in a new job. In Table A.23 we also document that individuals in the high-wage intervention are significantly less likely to have obtained a job in their desired occupation, a job that they see themselves doing in the long run or a job with an open-ended contract. One possible explanation for these results is that the individuals that are induced to apply to

[30] We are similarly unable to document systematic impacts on ability when we ipsatise the psychometric measures to correct for acquiescence bias (the tendency to agree with any statement, regardless of the content of that statement). We report these additional regressions in Table A.20.

[31] In this Table, we exclude the application to the experiment's job in the definition of the dependent variables. In Table A.22, on the other hand, we report estimates for the *total* number of applications, which includes the application to the experiment's job. Under this alternative definition, we estimate that both interventions are associated with a marginally insignificant increase of about .1 applications per jobseeker.

the experiment's job by the higher wage run out of resources to search for other jobs. The effects of the high-wage offer on search effort are indeed negative: the intervention is associated with a significant 10 percent decrease in the number of visits to the job vacancy board — a costly form of search that requires the use of public transport (Table A.24). In Table A.21 we also document effects on applications that point in the same direction, but are smaller and statistically insignificant: a 4 percent decline in the number of applications to other jobs, and a 3 percent decline in the time spent on job applications.

5.4 Mechanisms: Who drives the increase in quality?

We study the heterogeneity of treatment effects along several dimensions. These include demographic characteristics (gender and age), labour market variables (employment status and work experience) and a variable capturing the net present value of the job to each jobseeker. We calibrate this variable by comparing the control wage of the experiment's job to a forecast of the wage that each individual can expect to be paid in the market, based on their observables. We describe in detail the calibration procedure in Appedix A.6. For each dimension of heterogeneity x, we estimate a model of this form:

$$\begin{aligned} \mathtt{y}_i = &\beta_0 + \beta_1 \cdot \mathtt{incentive}_i \cdot I(x_i = 1) + \beta_2 \cdot \mathtt{high\ wage}_i \cdot I(x_i = 1) \\ &+ \beta_3 \cdot \mathtt{incentive}_i \cdot I(x_i = 0) + \beta_4 \cdot \mathtt{high\ wage}_i \cdot I(x_i = 0) \\ &+ I(x_i = 1) + \kappa_b + u_i. \end{aligned} \tag{8}$$

Model (8) gives us separate estimates of the effect of treatment for individuals for whom $x = 1$ and individuals for whom $x = 0$. When a variable is continuous, x is dummy that splits the sample at the median of that variable. For each regression and each treatment, we present a test of the hypothesis that there is no heterogeneity in the effect of that treatment ($H_0 : \beta_1 = \beta_3$ for the incentive, and $H_0 : \beta_2 = \beta_4$ for the high-wage offer). We bootstrap the standard errors of the regressions studying heterogeneity with respect to the value of the job, to reflect the fact that we are using a generated regressor. Results are reported in Tables A.25 to A.33 in the Appendix.

We find that the increase in cognitive ability caused by the incentive is significantly stronger among women, the unemployed, the less experienced, and for those individuals whom we estimate to value the job the most (Table A.27). Impacts are also much larger for applicants below the median age, though in this case we cannot reject the null of not heterogeneity ($p = 0.161$). These are all groups that on average fare worse in the labour market and that respond more strongly to job-search support (Card et al., 2010;

Abebe et al., 2020). Further, with the exception of work experience, we cannot document heterogenous impacts of the high-wage offer with respect to these dimensions. The magnitude of the differences in impacts across groups is often large. For example, among males, the effect of the incentive on average cognitive ability is close to zero. Among women, on the other hand, the cognitive ability score more than doubles (and the Raven test score increases by about 4.7 points or about .4 of a standard deviation, as reported in Table A.28). We also document significantly larger effects for women at the 90th and 75th percentiles. We illustrate these results graphically in Figure A.8, where we show that the proportion of female top applicants grows from 18 percent in the control group to 31 percent in the application incentive group.

On the other hand, we do not find systematic heterogeneity in the treatment effect on the probability of applying for the position (Table A.25). For each dimension of heterogeneity, we can use the estimated impacts on applications rates and group-specific quality to break-down the total treatment effect in three parts: a compositional effect, a within-group effect for the first group, and a within-group effect for the second group. This exercise, which we report in Table A.32, shows that the impact of the intervention is largely driven by within-group improvements in quality among those groups that have weaker employment prospects (women, youth, the unemployed, the inexperienced). This suggests that the effectiveness of the intervention could be improved by targeting the subsidy towards these groups, a point which we explore in depth in Section 6. Further, when we use all observables to obtain a forecast of market wages and then calculate the present value of the job, we find that virtually all of the increase in ability is driven by individuals with a low forecasted market wage and hence a high present value of the job. This finding helps us rule out a mechanism whereby the incentive increases ability by attracting individuals with strong outside options and hence a relatively modest personal gain from getting the job. On the contrary, it suggests that the application subsidy attracts high-ability individuals with weak outside options and thus has the potential to increase allocative efficiency.

5.5 Alternative explanations

In this section we consider seven alternative explanations for our results that are unrelated to application costs. We do not find evidence suggesting that these channels drive the effect of the application incentive.

Do the interventions change test effort? First, we consider whether the treatments affect test effort. This could happen for a number of reasons. For example, applicants in the wage intervention may exert more effort in the selection test since the higher salary makes the position more valuable. We study this alternative hypothesis in two ways. First, we note that the interventions have significant treatment effects on GPA (reported in Table A.12), a measure of ability that was determined before treatment and that hence cannot be confounded by differential test effort. Second, we follow Dal Bó et al. (2013) and collect a direct measure of test effort through a task that requires effort, but very little ability. In this task, applicants have to transcribe ten strings of meaningless letters. In Table A.36 we study whether two measures of performance on this task — the number of strings that have not been transcribed and the number strings that have been transcribed incorrectly — improve with treatment. We find that the average applicant in the control group fails to transcribe 0.08 out of 10 strings and transcribes incorrectly 0.7 out of 10 strings. Applicants in both treatment groups perform in a similar and statistically indistinguishable way to control applicants, both in terms of of mistakes and in terms of number of strings transcribed. For example, the average applicant from the high-wage intervention makes an additional .067 mistakes — a very small difference. We obtain similar results if we study the likelihood of making any mistake or failing to transcribe any string. Altogether, this evidence suggests that the treatments do not change test effort.

Do individuals in the incentive group apply primarily to collect the monetary payment? Next, we explore whether application incentives attract individuals that are primarily interested in the immediate monetary payment, and not in the position. We consider three implications of this hypothesis. First, if this hypothesis was correct, we would expect lower test effort among treated applicants. However, as discussed above, applicant test effort is very similar across experimental groups. In Figure A.13 we further show that the distribution of test effort in the control and incentive groups looks remarkably similar at all point of the support (and a Kolmogorov-Smirnov test fails to reject the equality of these distributions). Second, we would expect that, compared to the high-wage condition, incentives would generate an increase in application rates that is more skewed towards low GPA jobseekers, who have a low probability of securing the position and should hence discount the wage increase more heavily. This is the opposite of what we observe when we plot application rates against GPA (Figure A.6). Third, we would expect that applicants from the incentive group would be less interested in formal employment. We shed light on this by comparing how applicants in the various groups search for other jobs in the period between the two phone surveys (Table A.37). We find no meaningful differences in either number of applications (consistent with the results for the full sample of jobseekers reported in Table A.21) or in the probability of rejecting a job offer.

Does the application incentive help present-biased individuals commit? Third, we explore the hypothesis that application incentives help present-biased individuals follow through on their intention of applying to the experiment's job. On the day the application is due, a present-biased individual who plans to take the test may be tempted to deviate from his or her plan. The incentive makes it more costly for them to change their mind — as they would forgo both the future opportunity of getting the job *and* the immediate monetary payment — and can thus help them act in a more time-consistent manner. While plausible, this explanation is not supported by the data. We do not find evidence suggesting that the incentive attracts applicants who have different time preferences, which we measure through an incentivised task similar to the task proposed by Augenblick et al. (2015) (described in detail in Appendix A.3). In Table A.34, we report structural estimates of *average* present bias, discounting and cost of effort, which are obtained by pooling all individual decisions in the incentivised task. We find that, on average, individuals have a value of the present-bias parameter β that is less than one, indicating time consistency. This is very similar across treatment groups. In the same table, we further show that the discounting and cost-of-effort parameters of applicants from the incentive group are similar and statistically indistinguishable from the controls. The high-wage treatment, on the other hand, attracts significantly more patient applicants. This is not surprising, as the present value of the high-wage offer is greater for patient individuals. In Table A.35, we then present additional regression results for *individual* measures of present bias, obtained by running a separate estimation for each individual. About 30 percent of the population is present biased according to this individual measure. This share goes up by an insignificant 2 percentage points among application incentive applicants. Thus, overall, the evidence does not support the hypothesis that the effect of the incentive is driven by present biased individuals. In Table A.35, we also show that there are no significant differences in other economic preferences (e.g. risk and social preferences), which we measure through non-incentivised questions.

Do the interventions make the job more salient? Fourth, we study whether the treatments induce jobseekers to pay more attention to the experiment's job. In our context, individuals may suffer from stress and cognitive load. As a result, they may be inattentive, especially when the benefits of an action accrue in the future, and may thus forget to apply to some of the jobs they are interested in (Mullainathan and Shafir, 2013). The interventions could induce jobseekers to pay more attention to the experiment's job because they increase the cost of forgetting, or because they make the job stand out among others (Gabaix and Laibson, 2005; DellaVigna, 2009). This increased salience would reduce the likelihood that individuals forget to apply to the position, potentially driving some of the impacts on application rates that we have documented.

We study this alternative explanation in two ways. First, we note that a mechanism of this type is likely to work against the direction of our findings on ability, as cognitive load temporarily decreases cognitive ability (Mani et al., 2013). Second, we directly test this explanation by leveraging the fact that salient information is more likely to be remembered (Botta et al., 2010; Santangelo and Macaluso, 2013). If the interventions make the position more salient, we would expect treated jobseekers to remember more accurately the information about the job that was given to them. We implement this test in Table A.38, using a question in the second phone call that asks jobseekers to recall the wage that was discussed in the first phone interview. In the control group, we find that about 70 percent of individuals report the correct wage. The remaining subjects either report an incorrect figure, or declare that they do not remember. The average report has an absolute mistake of 167 ETB. Importantly, we are unable to find systematic evidence of better recollections for the application incentive group. However, we find that individuals in the high-wage group recollect the wage of the position more accurately: they are more likely to report the correct figure (by 3.8 percentage points), and they make smaller absolute mistakes (by 45 ETB on average). These results are robust to controlling for whether individuals have applied for the job.[32] Thus, overall, while these results suggest that the wage intervention increases salience, the evidence does not support the hypothesis that this is also the case for the incentive intervention.

Do the interventions change jobseekers' beliefs about their prospects in the labour market? We study whether individuals update their beliefs about their prospects in the labour market. This could be the result of a revision in the beliefs that individuals hold about their own employability, or in the beliefs about the state of the labour market. For this test we use two questions from the second phone interview. In the first question,

[32]Applicants have a second chance to inquire about the wage during the application process and may thus have better recollections for reasons unrelated to salience. Since the treatments increase application rates, this may bias upwards the coefficients in columns (1) and (2) of Table A.38. To correct for this, the models reported in columns (3) and (4) control for whether the individual has applied for the job. This control variable captures an endogenous mediator and hence the models in columns (3) and (4) should be interpreted as a basic form of mediation analysis (Acharya et al., 2016). We present a formal mediation analysis at the end of this section.

we ask subjects to forecast the number of weeks that it would take them to find a job that paid at least their reservation wage. In the second question, we ask respondents to report the wage that they expect this job will pay.[33] We find that the application incentive does not have a significant effect on either of these beliefs. The high-wage offer, on the other hand, significantly increases expected wages by about 8.5 percent. Table A.39 in the Appendix reports these results.

We also do not find evidence that the interventions change jobseekers' beliefs about the *probability of getting the experiment's job*. We show this result in Table A.40 in the Appendix. This finding is consistent with the low levels of strategic sophistication that are documented in a simplified beauty contest task (Crawford et al., 2013). In general, applicants are overconfident about their likelihood of getting the experiment's job. This is consistent with recent research showing that beliefs about individual performance are characterised by overconfidence in several contexts, including job search (Malmendier and Tate, 2015; Spinnewijn, 2015; Banerjee and Sequeira, 2020).

The impacts on beliefs do not vary significantly according to the ability of the jobseeker (as proxied by their GPA score) or previous work experience. We show this in Figures A.14 and A.15 in the Online Appendix, where we report impact estimates for various subgroups and a p-value for a test of the null hypothesis of no heterogeneity.

Do the interventions change jobseekers' beliefs about the attributes of the job? We test whether the treatments affect the beliefs that individuals hold about the characteristics of the experiment's job. To test for this, in the second phone call we collect jobseekers' beliefs about several attributes of the job: holidays, non-standard working hours, the degree of autonomy, how satisfying the work will be, whether they will learn new skills, etc... We regress each of these beliefs on the two treatment dummies and report results in Table A.41 in the Appendix. We find that the application incentive has a modest significant effect on two of these dimensions: the proportion of people who think the job will have more than four days of holidays per month goes up by 2 percentage points, and the proportion of people who think that the job will help them

[33] To elicit expectations about the wage, we follow the method of Attanasio and Kaufmann (2009). We ask respondents to report the minimum and maximum wage that the job can pay. We then identify the midpoint between these two values and ask respondents to report the probability that the job will pay more than the midpoint. Following Attanasio and Kaufmann (2009), we assume that beliefs follow a triangular distribution. This distribution is fully characterised by an upper bound, a lower bound and a mode. The maximum and minimum wage reported by respondents identify the upper and lower bounds. Given the two bounds, the value of the CDF at the midpoint identifies the mode of the distribution.

to find a job in the future goes up by 3 percentage points. These two expectations are weak predictors of the decision to apply for the experiment's job. Among control group individuals, the belief that the job has long holidays raises the probability of making an application by 7.8 percentage points, while the belief that this job will help with job search in the future raises the probability of making an application by 8.2 percentage points. To assess the potential effect of this channel on application rates, we multiply the treatment effects on the beliefs by the effects that these beliefs have on application rates and add up. The result is that this channel can explain a change in application rates of about half a percentage point. In other words, net of this effect, the application incentives would raise applications by 11 percentage points (as opposed to 11.5 percentage points).

We provide additional evidence that application incentives do not change beliefs about job attributes with a survey experiment on a new sample of jobseekers. We select 724 jobseekers through random visits at the same job-vacancy boards where the experiment's job was originally advertised.[34] During the survey, we give subjects some information about a hypothetical job, including salary and nature of the job, randomising whether this description mentions a monetary application incentive or not. We then ask respondents whether they expect the job to have a number of attributes (since the questions refer to a hypothetical job, respondents are not rewarded for correct answers). In Table A.43 we show that, in this new sample, application incentives do not generate any significant or sizeable change in job-attribute beliefs. Together with the results on retrospective beliefs reported in Table A.41, this evidence reassures us against the possibility that the impacts of application incentive are driven by a misperception of job attributes.[35]

[34]We carried out this survey between December 2019 and January 2020, as part of a different project. We report basic descriptives for this new sample in Table A.42. The jobseekers in this sample have broadly similar characteristics to the jobseekers in our original experiment: they are 24 years old on average, 20 percent of them were born in the capital city, about 40 percent have work experience and, those who do, have worked for 20 months on average. The corresponding figures for our original sample are: 26 years of age, 24 percent born in Addis Ababa, 53 percent have work experience and, those who do, have worked for 28 months on average. The two samples however differ in terms of gender: 49 percent of respondents in the new sample are female, as the sample was stratified by gender, while only 21 percent of the respondents in the original experiment are female. Table A.42 also reports a standard battery of balance tests. The overall test of orthogonality confirms that treatment assignment is balanced. However, we find some imbalance (p=0.08) in terms of the proportion of jobseekers who are currently unemployed.

[35]As an additional check, we debriefed the staff from the original experiment to find out whether treated applicants asked more or different questions about the position during the first phone call. The staff was unable to recollect any systematic differences in the questions asked by treated and control subjects.

Mediation analysis. We use mediation analysis to quantify the contribution of the changes in beliefs and salience caused by the wage intervention to its overall impact on application rates. We do this by estimating the *average controlled direct effect* (ACDE) of this intervention (Vansteelandt, 2009; Acharya et al., 2016). This quantity is defined in a framework where a treatment can have both a direct impact on the outcome of interest and an indirect impact which runs through a mediator. For example, the wage intervention may have a direct impact on application rates as it makes the position more lucrative, but also an indirect effect if it makes potential applicants more confident. The ACDE is defined as the direct effect of the intervention, that is, the effect that the intervention would have if the mediator was not allowed to respond to treatment and hence the indirect effect was removed.[36] Here, we focus on the ACDE on application rates and consider a list of potential mediators that we showed to be affected by the interventions in the analysis above: beliefs about labour market prospects, beliefs about the attributes of the experiment's job, salience of the position.

We find that the ACDE of the high-wage offer on application rates is 9 percentage points (with a 95% confidence interval ranging from 3 to 13 percentage points). This is significantly smaller than the original estimate reported in Table 2 (which was 18.7 percentage points, with a 95% confidence interval ranging from 15 to 22 percentage points), suggesting that the mediators have a quantitatively large influence on application rates in the high-wage group. The controlled direct effect of the incentive, on the other hand, is quantitatively similar and statistically indistinguishable from the original treatment effect (the demediated effect is 13 percentage points, slightly larger than the original effect of 11.5 points). This is not surprising, as we only find evidence of large and systematic effects on the mediators for the wage treatment.

5.6 Other characteristics of applicants

We report one final set of results on the characteristics of applicants. We are particularly interested to shed light on a set of productivity-relevant traits that may not be captured

[36] In order to identify the ACDE we have to assume *sequential unconfoundedness*. In a case where treatment is randomly assigned, this amounts to assuming that there are no omitted variables which confound the effect of the mediator on the outcome, conditional on treatment and a set of pre-treatment controls (Acharya et al., 2016). Given this assumption, we can identify the ACDE with a simple two-step procedure. In the first step, we regress the outcome on the mediator, the treatment dummies, a set of controls, and the interaction between the mediator and all other variables. We then obtain the predicted value of the outcome fixing all mediators to zero. This is the 'demediated' outcome. In the second step, we regress the demediated outcome on the treatment dummies. The coefficients from this regression give us the estimate of the ACDE.

by our main indicators of ability. These include confidence and motivation, and personal constraints that may affect the ability to work. In Table A.44, we report regressions for a comprehensive battery of psychometric measures — grit, locus of control, core self-evaluation, self-esteem as well as the individual big-five traits. We are unable to find any meaningful differences between control and incentive applicants: all effect sizes are markedly below .1 standard deviation and are precisely estimated. Further, in Table A.34 and Table A.35 we show that treated applicants have similar economic preferences and a similar cost of effort as control applicants. As explained above, our measure of cost of effort is obtained in a real-effort task where individuals have to allocate a certain amount of work across two different work sessions. In this task, treated applicants make similar allocation decisions as control applicants — a result that suggests that they have a similar ability to schedule work in advance and to work without interruption as control applicants. Overall, the results of this further analysis suggest that, apart from the differences in cognitive ability highlighted above, the interventions attract pools of applicants that have characteristics similar to those of control applicants.

6 Structural analysis

In this section we discuss the identification and estimation of the structural model. We then present and interpret the estimates of the structural parameters. We find that application costs are large, heterogeneous and positively related with ability. These results are consistent across a number of robustness tests.

6.1 Identification and estimation

Our objective is to estimate the parameters that characterise the joint distribution of costs (C) and ability (T), for each level of the value of the job (B). In our main empirical model, we allow for two possible values of B. Thus, we need to identify ten parameters for the joint distribution of T and C of low and high-B jobseekers. We also estimate the magnitude of the shocks to costs and benefits of the two interventions (τ and τ^w) and the parameters that capture perceived selectivity. In the case where ability is noisy, we estimate a single selectivity parameter a (which may differ from true selectivity if jobseekers have an inaccurate perception of the selection process). In the noisy selection case, we assume that jobseekers believe selectivity is an independent, normally distributed variable with mean μ_a and standard deviation σ_a. Thus, we have a total of thirteen parameters to estimate in the noisy-ability case, and fourteen parameters in noisy-selection case.[37]

To identify these structural parameters we use a core set of fourteen empirical moments, which we obtain from administrative data on applications and test scores. We then show that our results are robust to the introduction of two additional moments based on self-reported beliefs data, which bring additional information on jobseeker uncertainty. For the core set of moments, we first use the control application rate and the control average and standard deviation of applicant ability (three moments). Further, we use the treatment effects on application rates and on average applicant ability of the two interventions (four moments). For the high-wage group, we use the demediated treatment effect on application rates, which nets out the indirect impacts of this intervention on applications through changes in beliefs and salience (see Section 5.5 for a full discussion of demediation). We compute these moments separately for low and high-B jobseekers, giving us a total of fourteen core moments. Jobseekers' average belief about the probability of being offered the job, for the low and high-B group, give us the two additional moments.[38] We present analytical formulas for all moments in Appendix A.4.

The parameters are jointly identified by these empirical moments. The basic intuition for identification is as follows. The six control group moments — the average and standard deviation of applicant ability and the application rate — describe the truncated distribution of T among applicants. These moments enable us to identify the mean of application costs (μ_C) — a key driver of the application decisions — and, conditional on this, the mean and standard deviations of ability (μ_T, σ_T) among all jobseekers. Then, the treatment effects on application rates identify the severity of the shocks τ and τ^w and the standard deviations of costs σ_{Ch} and σ_{Cl}. Further, the treatment effects on applicant ability identify the covariance between cost and ability, as shown in Proposition 2 and Proposition 3. Finally, perceived selectivity is identified by the treatment effects on ability and by jobseekers' forecast of the probability of being offered the job. Tables A.46 and A.47 summarise these intuitions.

We proxy ability T with the score on the Raven test and then study whether our findings change if we use the cognitive ability score instead. Further, to measure B

[37]For noisy-ability case, these parameters are: μ_{Tl}, μ_{Cl}, σ_{Tl}, σ_{Cl}, σ_{TCl}, μ_{Th}, μ_{Ch}, σ_{Th}, σ_{Ch}, σ_{TCh}, τ, τ^w and a. For the noisy-selection case, these parameters are: μ_{Tl}, μ_{Cl}, σ_{Tl}, σ_{Cl}, σ_{TCl}, μ_{Th}, μ_{Ch}, σ_{Th}, σ_{Ch}, σ_{TCh}, τ, τ^w, μ_a and σ_a.

[38]We elicit beliefs about the probability of being offered the experiment's job retrospectively. During the second phone call, we ask subjects to report the belief that they held at the time of deciding whether to apply to the position or not (that is, shortly after the first phone call). We show that these beliefs are not affected by the interventions in Table A.40 in the Online Appendix.

— the value of the experiment's job net of outside options — we first obtain a post-LASSO forecast of the wage a jobseeker can expect to earn given their observables. We then calibrate B using the wage paid by the experiment's job in the control condition, and an informed assumption about the probability of finding an alternative job at the jobseeker's individual market wage. We discuss this calibration in detail in Appendix A.6. For the estimation, we discard individuals with a negative B since our model is only specified for the case where $B > 0$. We split the remaining individuals (about 61 percent of the sample) at the median level of B. On average, an individual in the high-B group gets a net, discounted benefit from the experiment's job of about 594 ETB (27.3 USD). For the low-B group, the benefit is about 407 ETB (18.7 USD). A possible drawback of this strategy is that the residual variation in the value of the job that is not captured by the empirical measure of B may inflate our estimates of application costs and of the cost-ability correlation. We explore this point empirically in Section 6.4, by studying the robustness of our key results to estimation strategies that leverage additional variation in B.[39]

To estimate the model we use a classical minimum distance estimator (Wooldridge, 2010). We save the empirical moments in a vector $\boldsymbol{m}$. For a parameter vector $\boldsymbol{\theta}$, we solve the model and calculate the simulated moments $\boldsymbol{m}_S(\boldsymbol{\theta})$. We update $\boldsymbol{\theta}$ in order to solve:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \left[\boldsymbol{m}_S(\boldsymbol{\theta}) - \boldsymbol{m}\right]' \cdot J(\boldsymbol{m})^{-1} \cdot \left[\boldsymbol{m}_S(\boldsymbol{\theta}) - \boldsymbol{m}\right], \tag{9}$$

$J(\boldsymbol{m})$ is a diagonal matrix that contains the variance of each moment, ensuring that more precisely estimated moments get a greater weight in estimation. In line with the recent literature, we use this simple weighting matrix instead of the theoretically optimal weighting matrix, which may suffer from small sample bias (Altonji and Segal, 1996). We calculate $J(\boldsymbol{m})$ using a bootstrap with 1,000 replications. We perform inference using a bootstrap (keeping $J(\boldsymbol{m})$ fixed). We include the estimation of B and the demediation procedure in all bootstrap exercises.

Table 4 presents our main structural results. We include two sets of estimates for the noisy-ability case, obtained with the core set of moments and with the extended

[39] These strategies capture variation in B that is generated by market wages. On the other hand, we are unable to capture variation in B that comes from heterogeneity in the value of non-work time, as we do not have individual-level data on the likely determinants of this variable (e.g. childcare, access to government welfare programs, etc.). An additional point to note is that our interventions may have an impact on jobseekers' beliefs about market wages. In Table A.39, we presented evidence suggesting this may be the case for the wage intervention, but not for the application incentive intervention. This means that the shock τ^w that we estimate captures the net effect of two countervailing forces: a direct impact on the value of the experiment's job (which raises the value of B) and an indirect impact on the value of jobseekers' perceived outside options (which lowers the value of B).

set of moments. We also include two sets of estimates for the noisy-section case, again obtained with the two different sets of moments. In what follows, we first describe the empirical fit of these four versions of the model, and then discuss the parameter estimates and their robustness.

< Table 4 here. >

6.2 Model fit

We generally obtain a good fit between empirical and simulated moments. We show this in Figure 7 below (and report the underlying moments in Tables A.49-A.52 in the online Appendix). In both the noisy-ability and the noisy-selectivity cases, the simulated moments fit the empirical moments tightly. Simulated application rates are generally within one percentage point of the true empirical moment. The mean and standard deviation of the Raven test in the control group are matched almost exactly. Finally, we also fit fairly precisely the treatment effect on applicant ability. Overall, the noisy-ability and noise-selection cases generate similar goodness-of-fit statistics, which we report in the last row of Table 4.

< Figure 7 here. >

Further, the two versions of the model that do not target beliefs are able to reproduce jobseekers' substantial overconfidence about the probability of being offered the job. Among individuals with $B > 0$, the average forecast of this probability is 47 percent. This is outside a reasonable range, since the employer hires one person approximately every 115 applicants. Importantly, when we do not the explicitly target these beliefs, our estimates imply a similar high degree of overconfidence. When ability is known, the implied average forecast of the probability of being offered the job is 64 percent. When ability is noisy, the average forecast is 35 percent. In addition, when we explicitly target these beliefs, both models match them very closely.

The estimated model also replicates a key non-targeted pattern in the data: in the control group, applicants have on average lower ability than non-applicants (as shown in Figure A.5 for GPA, the only measure of ability that we observe for all jobseekers). This pattern intuitively suggests that the marginal applicant is better than the average applicant and hence, that the employer can attract better applicants by intervening on either the cost or the benefit margin. The estimated model reproduces this key pattern for both the noisy-ability and the noisy-selection case.

Finally, the elasticity of the simulated moments with respect to the model parameters broadly supports the intuition for identification presented above. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the estimated parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments. This illustrates the drivers of identification close to the minimum of the estimation problem. We report the results in Tables A.53-A.56 in the Appendix. The estimated elasticities are generally consistent with the intuitions for identification discussed above. For example, the elasticity of the treatment effect of application incentives on applicant ability with respect to the cost-ability covariance is 3.8 (a 1 percent change in the covariance leads to a 3.8 percent change in the simulated moment). For the moments that describe application rates or ability in the control group, the elasticity with respect to this parameter is much lower. Tables A.54 and A.56 also illustrate how jobseekers' beliefs help identify perceived selectivity: these moments are most responsive to changes in the selectivity parameters.

6.3 Parameter estimates

6.3.1 Noisy-ability case

Our estimates for the noisy-ability case indicate that application costs are large, heterogeneous, and positively correlated with ability. We report these estimates in the first two columns of Table 4. Column (1) reports the results obtained using the core set of empirical moments. For the high-value group, the mean of application costs is 217 ETB. This amounts to 13.5 percent of the monthly salary offered to individuals in the control group, or to about 36.5 percent of the estimated value of the job. For the low-value group, mean costs are about 136 ETB, or 8.5 percent of the salary and 33.5 percent of the value of the job. We also estimate that application costs have a large dispersion, in both groups. The standard deviation of application costs is about 245 ETB for the high B group and 193 for the low B group.

The correlation between ability and costs is about 0.57 for the high B group and 0.64 for the low B group. These estimates imply a large increase in ability as we move along the cost distribution. For example, a high-B jobseeker with costs one standard deviation above the mean has a Raven score that is about 6.4 scores higher than the average jobseeker (a 15 percent increase). Using the average Mincerian return to cognitive ability reported in Section 2, we estimate that the value of this additional ability would be 146 ETB per month.

When we introduce jobseekers' beliefs, we obtain similar estimates of the cost-ability correlation and somewhat higher estimates of average costs (column (2) of Table 4). The estimated selectivity threshold is lower, which helps fit jobseekers' beliefs. All other parameters estimates are qualitatively unchanged.

6.3.2 Noisy-selection case

Our estimates for the noisy-selection case confirm that the magnitude of application costs and of the cost-ability correlation is substantial. We report these estimates in the last two columns of Table 4. We estimate correlation coefficients that are consistent with those of the noisy-ability case. However, mean application costs are substantially larger. When we use the core set of moments, we estimate costs of 273 ETB and 383 ETB. When we bring in the additional moments, we estimate costs of 206 and 286 ETB.

Uncertainty about selectivity is very large. As discussed in Section 3, this uncertainty moderates the size of the "standard intuition" effect. We can show this by computing the increase in average applicant ability that we would observe if the correlation between cost and ability was set to zero (and all other parameters were equal to those reported in column (3) of Table 4). Thus, in this simulation, the treatment effect on ability is driven entirely by the "standard intuition" effect. We calculate that, for the high B group the drop in average ability would correspond to 0.002 correct answers on the Raven test. This is less than 1 percent of the overall positive effect on ability that we simulate if we use the estimated positive value of the correlation. Thus, these findings suggest that the screening role of application costs is limited in our context.

6.3.3 Plausibility checks

To explore the plausibility of our estimates of application costs, we leverage jobseekers' own reports about the monetary and time costs of other job applications.[40] The average monetary expenditure for one application is 50 ETB.[41] Further, jobseekers report that the average time required for one application is about 5 hours — which is broadly in line with the time required to apply for the experiment's position. We estimate that the

[40]This data was collected during the second phone interview. The question we asked referred to applications to positions other than the experiment's job, in the 30 days between the two phone interviews. While most jobseekers make at most one job application in this period, a small group sends a large a number of applications. This last group also spends much less money and time per application, which complicates the interpretation of these cost values. Thus, in the analysis that follows, we use the average figures reported by jobseekers that make only one application.

[41]This figure is likely to reflect the multiple trips jobseekers make to complete the application process (in-person application, written test, interview) as well as printing CVs and other one-off expenses. It is thus not an estimate of the cost of a single day of commuting.

value of this time is between 56 ETB and 89 ETB.[42] The sum of monetary and time costs is thus between 106 and 139 ETB. This should be seen as a lower bound to the total cost of the application since it does not include non-material costs (related to stress, attention, effort, etc.). Given this, our structural estimates for the noisy-ability case (136 and 217 ETB, with the core moments) are broadly consistent with these figures. On the other hand, the estimates for the noisy-selection case are noticeably larger.

Second, to assess the plausibility of our correlation estimates, we run a simple simulation based on the selection patterns we documented using the high-frequency panel. This panel tracks one cohort of jobseekers over the course of one year. For our simulation, we assume that the labour market is composed of multiple cohorts that live for three years, shrinking in size by 20 percent per year. Further, we assume that the cost-ability correlation in each cohort evolves as in the cohort that we observe in the panel data. Given this set-up, we simulate a steady-state correlation between cost and ability of about .35, which confirms that a substantial correlation is plausible in this labour market. However, this benchmark is lower than our structural estimates of the correlation parameter (which are between 0.57 and 0.64). This may be due to the fact that our proxies of application costs are noisy, which would bias the simulation downwards. Alternatively, our structural estimates may be inflated by variation in the value of the job that is not captured by our empirical measure of B. We explore this point in the robustness section below.

6.4 Robustness

The results on application costs and their correlation with ability are robust to the use of several alternative estimation strategies. For this robustness analysis, we focus on the noisy-ability case, which gives the most conservative estimates. We perform the following checks: (i) we use the cognitive ability score as opposed to the Raven score to measure ability, (ii) we let, in turn, a, τ^w, and τ differ by B group, (iii) we predict B using an OLS model instead of the post-LASSO estimator and (iv) we allow for three

[42]The opportunity cost of time is high for people in our sample, who may not have a formal, stable job, but often have access to casual, informal income-generating opportunities. These opportunities are typically as remunerative, on an hourly wage basis, as entry-level formal jobs in the bottom part of the wage distribution. Thus, if we value the time spent on the application at the rate of the salary offered in the low wage condition of our experiment (assuming a 5-days, 40-hours working week and 2 statutory days off per month), the opportunity cost of time is about 56 ETB. This goes up to 89 ETB if we assume that a whole day worth of income-generating opportunities is lost — a reasonable assumption since one may have to forgo, say, a day of casual work at a construction site in order to spend 5 hours at the application centre.

types of B in the population. We report all key results in Table A.57.

First, we find that, when we use the cognitive ability score to measure ability, we estimate very similar levels of costs and a higher cost-ability correlation. These results, reported in the first column of Table A.57, suggest that our structural findings are not driven by our particular measure of ability. Second, we find that if we relax the assumption that a, τ, and τ^w do not vary across groups, we estimate values for the high and low-B groups that are qualitatively very similar to each other, as shown in columns (2)-(4). Third, we study two models designed to reduce the residual variation in B, either by splitting the distribution of B more finely, or by forecasting market wages using an OLS estimator (which extracts information from a larger set of covariates compared to the Post-LASSO estimator). When we split B into three groups, application costs range from 153 ETB to 274 ETB and the correlation is between 0.50 and 0.65. Further, when we use the OLS estimator, the lowest estimate of application costs is 115 ETB and the lowest estimate of the cost-ability correlation is 0.51. Thus, under both strategies, we estimate large application costs and a strong cost-ability correlation. While this exercise cannot fully rule out that our estimates may be inflated by residual variation in the value of the job, it provides evidence of the qualitative robustness of our findings when this residual variation is plausibly reduced.

6.5 The returns of the interventions and policy simulations

In this last subsection, we assess the returns of the interventions and of two counterfactual policies. Both interventions attract applicants with higher cognitive ability and thus enable the employer to make better hires. Since cognitive ability is a strong predictor of productivity and the wage is fixed to the level that was originally posted, this generates an expected stream of benefits for the employer. Each intervention also entails two types of costs. First, the employer has to pay the direct cost of the intervention (the cost of the incentive or the higher wage). Second, the employer has to employ staff time to review the additional applications.

We calibrate our cost and benefit calculations in order to assess the effect of the interventions on a single, typical employer recruiting a clerical worker in this market. For these calibrations, we use the data that we collected from local firm managers. First, we quantify recruitment costs using managers' assessment of the time required to review one more application. On average managers report that this requires about one hour of work.[43] We price this hour at the median salary of the HR staff who review applications

[43] We also ask whether there are any financial costs involved in reviewing one more application. The great majority of managers report that this is not the case. The majority of financial costs are fixed costs,

in these firms. Second, we calibrate the number of applicants in the control group and the number of jobs on offer using the average of these variables among the firms in our sample. Third, we compute worker turnover rates and use these to assess the expected number of months that the worker will spend in the firm (expected tenure is 45 months; we assume, conservatively, that the high wage is paid only for three months). Finally, we calibrate the productivity gains from higher worker ability using the Mincerian regressions on local labour market data that we presented in Section 2. We summarise these assumptions and give additional details on our calculations in Appendix A.4.

We design two counterfactual policies that reduce the upfront costs of the application incentive. One drawback of this intervention is that the employer subsidises a large group of infra-marginal individuals who would have applied for the job in the absence of the incentive. Further, while most of the increase in ability is driven by the high-B group, the incentive is offered to both groups. To decrease transfers to infra-marginal and to low-ability applicants we propose two alternative policies: (i) an application incentive that is targeted on the basis of observable demographics (we experiment, in turn, with offering the incentive only to women and only to individuals below the median age),[44] (ii) an application incentive that is offered only to the applicants who score above a threshold in the test (we set this threshold to the level that fills the positions on offer in expectation, so in practice under this policy the incentive is offered to all hires). These interventions reduce transfers to infra-marginal individuals by exploiting, in turn, the observable information available to employers and the information available to workers. One important caveat with respect to the second counterfactual intervention is that our estimates do not take risk aversion into account. Under this intervention, jobseekers are uncertain about whether they will qualify for the incentive. If risk aversion is large, the incentive will be less attractive than in our simulations. Further, if it is negatively correlated with ability, the positive effect on the quality of the pool of applicants will be less pronounced than what we predict. The returns that we estimate for this intervention should thus be seen as an upper bound of the true

such as advertisement costs.

[44]To estimate these counterfactuals, we estimate the structural model two more times, using new empirical moments obtained by splitting the sample by age and by gender. We report these new parameter estimates in Table A.58 and A.59, and the related moments in Table A.60 and A.61. The model that explores heterogeneity by age fits the data well. On the other hand, the model that explores heterogeneity by gender has a poorer fit. This is driven by the fact that, for the incentive intervention, the model predicts an increase in the ability of female applicants that is only about one fifth of the impact observed in the data. In our IRR calculations, we thus conservatively augment the simulated impact of this intervention, reducing the gap between the empirical and simulated moment by 25 percent. Our estimates are hence likely to be a lower bound of the true returns of this counterfactual intervention.

returns.

< Table 5 here. >

We present these results in Table 5 below. First, we find that the application incentive has a positive internal rate of return (IRR) of about 9.8 percent. This is above market interest rates (which were about 5 percent in Ethiopia at the time of the experiment), and in line with the hurdles rates commonly reported by firms.[45] However, the confidence interval of this intervention has a negative lower bound. Thus, while the intervention is attractive in expectation, managers may be concerned about its downside risk. Second, we find that all counterfactual incentive schemes have a large impact on the return of the intervention. Targeting the incentive to women or to jobseekers below the median age raises the IRR to 86 percent and 47 percent, respectively. Both of these schemes substantially reduce the cost of the intervention, as they restrict the number of applicants that are eligible for the incentive. At the same time, they generate large gains in applicant ability, since they target the incentive on the demographics that drive the impacts of the intervention. Thus, they raise the IRR considerably. Further, the lower bound of the confidence interval of both of these policies implies a positive return; targeting the incentive on selected demographics thus substantially reduces the downside risk of the original policy. Third, when the incentive is offered to all hires, the cost of the intervention decreases dramatically, and the IRR goes up to 382 percent. However, as discussed above, this large IRR should be interpreted as an upper bound, as it does not consider the role of risk aversion.

Our IRR estimates do not measure whether the intervention generates overall welfare gains. The fact that high-B individuals drive the impacts on ability (Table A.27) suggests that application incentives may have positive effects on the allocation of talent, favouring those high-ability jobseekers who stand to gain the most from being offered the job. However, we do not have access to the comprehensive labour market data that would be required to credibly quantify these allocative gains. In particular, we do not have information on how other firms are affected by the policy or on workers' long-term outcomes. We thus do not provide an estimate of welfare effects.

[45]We are not aware of data on the hurdle rates used by firms in Ethiopia or in countries with similar macroeconomic conditions. A recent survey by the Bank of England finds that most firms in the UK adopt hurdle rates between 5 and 15 percent (Saleheen et al., 2017).

7 Discussion

In this section we discuss two important questions that follow from our findings. First, we study whether employers value cognitive ability and whether they are aware of the impacts of application incentives on the quality of the applicant pool. Second, in order to shed light on the external validity of our results, we present evidence from other contexts on application costs and their correlation with applicant ability.

7.1 Do managers value cognitive ability and are they aware of the benefits of application incentives?

We address these questions through a second novel experiment that studies the preferences and expectations of managers at firms recruiting for clerical positions (see section 2 for a description of how these firms are sampled). This experiment enables us to explore two possible reasons why firms in Addis Ababa do not use application incentives. First, firm managers may not value general cognitive ability and thus may not rank applicants from the incentive group above control applicants. Alternatively, managers would like to recruit workers with higher cognitive ability, but do not expect that the application incentive will attract these workers.

In the first task, we study whether firm managers rank treatment group applicants above control applicants. To incentive this task, we offer to invite one person from our sample of applicants to make a new job application at the manager's firm. The manager can determine who this person will be by ranking the standardised CVs of three selected applicants.[46] We sample one applicant from each experimental group. At this point of the experiment, however, the manager has not been informed about the two interventions nor about how the three applicants have been selected. On the CVs, we report applicants' education, age, work experience, GPA and the results from the Raven and conscientiousness tests (Figure A.16 shows a sample CV). We select triplets of applicants that reproduce as closely as possible the average differences in these characteristics between groups.[47] After the manager ranks the CVs, we randomly draw two

[46] Kessler et al. (2019) propose a similar methodology — called Incentivised Resume Rating — to elicit employer preferences. There are two main difference between their methodology and ours: (i) they use fictitious CVs while we rely on real CVs, and (ii) in their design, employers assess CVs by reporting a cardinal score, while in our design employers report a rank.

[47] We do this in three steps. First, we generate a list of all possible triplets. Second, for each triplet, we calculate the distance between (i) the differences in ability among the applicants in the triplet and (ii) the average differences in ability among their respective experimental groups (with a simple sum-of-square-differences statistic and three measures of ability: GPA, Raven score and conscientiousness). Third, we

of the three CVs and invite the person with the higher rank to make an application at the manager's firm. This last feature ensures that the manager is incentivised to report truthfully her preferences over the three candidates.

We find that both interventions improve the quality of the applicant pool in the eyes of local firm managers. We show this result in Table 6 using a series of linear probability models. In the first two columns, the dependent variable is a dummy for individuals who are ranked first. In the third column, the dependent variable is a dummy for being ranked first or second. We find that applicants from the incentive group are a significant 36.9 percent more likely to be ranked first than control applicants, and a significant 37.4 percentage points more likely to be ranked first or second. In column two, we only consider applicants from the control and incentive groups. We find that incentive group applicants are ranked above control group applicants about 70 percent of the times.

< **Table 6 here.** >

In the second task, we test whether managers can predict the effects of the application incentive. We first give managers detailed information on the experiment and then ask them to forecast the impacts of the application incentive on application rates and applicant ability (as measured by the Raven test). To measure ability at different points of the distribution, we obtain forecasts of (i) the average Raven score and (ii) the average Raven score among the 100 highest-scoring applicants. Further, before forecasts are made, we disclose the application rates and Raven test scores of applicants in the control and high wage groups in order to anchor managers' priors on the correct level of these variables. We reward managers for the accuracy of one randomly drawn forecast.

We find that managers make considerable forecasting errors and generally underestimate the impacts of the application incentive on applicant ability. In Figure 8 we plot the distribution of managers' forecasts. On average, managers expect that the application incentive will increase application rates and decrease applicant ability. In particular, they predict that performance on the Raven test will fall by about one correct answer, both at the mean and at the top of the distribution. In reality, performance on the Raven test improves by about one correct answer in both cases. Overall, about 66 percent of managers underestimate the level of cognitive ability of the applicants in the incentive group.

choose the sixteen triplets that have the minimum value of this distance statistic. We then randomly allocate a triplet to each manager. Across triplets, we also randomly change the order with which the candidates from the three experimental groups are presented.

In sum, the evidence in this section shows that managers value applicant cognitive ability, but are unaware of the cognitive ability gains generated by application incentives.

< **Figure 8 here.** >

7.2 External validity

We find several pieces of evidence suggesting that job search and applications costs are large in a number of contexts. In China, Chang (2009) reports qualitative evidence that manufacturing workers do not apply for attractive jobs when the application centre is too far from their place of residence. In Ghana, Hardy and McCasland (2017) provide evidence that entry fees for apprentices create labor constraints for small firms. In South Africa, Abel et al. (2017) show that jobseekers apply to fewer jobs per week than they would like to. A psychological intervention is only able to reduce this intention-behaviour gap by a small amount, suggesting that other constraints such as credit constraints may be responsible for jobseekers' failure to meet their job application targets. In OECD countries, employers often reimburse applicants for travel expenses incurred for job interviews. In Germany, this is mandated by the Civil Code.[48] In other markets, this emerges as a stable equilibrium outcome. An example of this is the economics job market, where most universities reimburse applicants for 'flyout' expenses. In all of these contexts, job application costs emerge as a salient constraint for jobseekers.

We also find observational evidence suggesting that these costs may be positively correlated with worker ability in several settings. In India, Choudhury and Khanna (2014) leverage micro-data from a large technology firm to show that hires from remote small towns (who face larger search costs for formal work) have higher ability and are more productive than hires from large cities (who face lower search costs). In the US, Paserman (2008) structurally estimates that, among unmarried jobseekers with low-wage work experience, high-ability jobseekers face search costs that are almost 5 times higher than those of low ability jobseekers.[49] Finally, there is evidence of similar selection dynamics in the context of college admissions. For example, Black et al. (2020) study the impacts of a policy granting public university access to all Texas students in the top 10 percent of their high-school class. They find that those top students from

[48]See §670 of the German Civil Code (BGB) and https://www.euraxess.de/germany/information-assistance/work-permit/faq-working.

[49]US jobseekers with low-wage employment experience are likely to be the most comparable group to the jobseekers in our sample. For jobseekers with experience in better-paid positions, Paserman (2008) finds a negative cost-ability correlation.

schools with poor university placement records who attend college thanks to the policy — a group that presumably faces large costs to access higher education — have better graduation rates than the students from the traditional "feeder" schools who are displaced by the policy. Overall, this evidence suggests that, across a variety of contexts, high-ability marginal applicants tend to face high application costs. In these settings, reducing application costs can improve both the efficiency and equity of the selection process.

8 Conclusion

In a worker recruitment experiment in Addis Ababa, Ethiopia we show that employers can use application incentives to attract applicants with higher cognitive ability. We estimate a structural model of applications decisions and find that the positive effect of application incentives follows from the fact that application costs are large, heterogeneous and, surprisingly, positively correlated with jobseeker ability. Finally, using a high-frequency panel dataset on job-search decisions, we present evidence suggesting that the correlation between application costs and jobseeker ability is likely to be driven by a dynamic selection mechanism: low-cost, high-ability jobseekers find work faster than high-cost jobseekers with similar ability and thus quit job search earlier.

Our experimental evidence on how application costs affect firms' ability to recruit talented workers generates a number of leads for future research. A first natural question is how incentives interact with interventions that improve the quality of screening (Autor and Scarborough, 2008). Second, our study does not investigate the potential dynamic gains from relaxing labour constraints. If personnel ability is complementary to capital and technology (Bender et al., 2016), these gains could be large. More broadly, our results suggest that unequal access to labour markets can distort the selection and allocation of talent. This central insight can help design future labour market policy.

References

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2020). Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City. *Working Paper*.

Abel, M., R. Burger, E. Carranza, and P. Piraino (2017). Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment.

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review 110*(3), 512–529.

Alatas, V., R. Purnamasari, M. Wai-Poi, A. Banerjee, B. A. Olken, and R. Hanna (2016). Self-Targeting: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy 124*(2), 371–427.

Alonso, R. (2018). Recruiting and Selecting for Fit. *Working Paper*.

Altonji, J. G. and L. M. Segal (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484).

Ashraf, N., O. Bandiera, S. S. Lee, et al. (2018). Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services. *Working Paper*.

Ashraf, N., J. Berry, and J. M. Shapiro (2010). Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia. *American Economic Review 100*(5), 2383–2413.

Attanasio, O. and K. Kaufmann (2009). Educational Choices, Subjective Expectations, and Credit Constraints. *NBER Working Paper No. 15087*.

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, 1067–1115.

Autor, D. H. and M. J. Handel (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. *Journal of Labor Economics 31*(S1), S59–S96.

Autor, D. H. and D. Scarborough (2008). Does Job Testing Harm Minority Workers? Evidence from Retail Establishments. *The Quarterly Journal of Economics 123*(1), 219–277.

Balakrishnan, U., J. Haushofer, and P. Jakiela (2015). How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments. *Working Paper*.

Bandiera, O., I. Barankay, and I. Rasul (2011). Field Experiments with Firms. *The Journal of Economic Perspectives 25*(3), 63–82.

Banerjee, A. V. and A. F. Newman (1993). Occupational Choice and the Process of Development. *Journal of political economy 101*(2), 274–298.

Banerjee, A. V. and S. Sequeira (2020). Spatial Mismatches and Imperfect Information in the Job Search.

Barrett, G. F. and S. G. Donald (2003). Consistent Tests for Stochastic Dominance. *Econometrica 71*(1), 71–104.

Beaman, L., N. Keleher, and J. Magruder (2013). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Working Paper*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *The Journal of Economic Perspectives 28*(2), 29–50.

Belot, M., P. Kircher, and P. Muller (2017). How Wage Announcements Affect Job Search Behaviour - A Field Experimental Investigation. *Working Paper*.

Bender, S., N. Bloom, D. Card, J. Van Reenen, and S. Wolter (2016). Management Practices, Workforce Selection and Productivity. *NBER Working Paper 22101*.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Black, S. E., J. T. Denning, and J. Rothstein (2020). Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes.

Botta, F., V. Santangelo, A. Raffone, J. Lupiáñez, and M. O. Belardinelli (2010). Exogenous and Endogenous Spatial Attention Effects on Visuospatial Working Memory. *The Quarterly Journal of Experimental Psychology 63*(8), 1590–1602.

Bruhn, M. and D. McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bryan, G. and M. Morten (2015). Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia. *Working Paper*.

Bureau of Labor Statistics (2017). News Release Job Openings and Labor Turnover: June 2017.

Card, D., J. Kluve, and A. Weber (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *The economic journal 120*(548).

Caselli, F., J. Ponticelli, and F. Rossi (2014). A New Dataset on Mincerian Returns. *Technology differences over space and time*, 108–118.

Chamorro-Premuzic, T. and A. Furnham (2010). *The Psychology of Personnel Selection*. Cambridge University Press.

Chang, L. T. (2009). *Factory Girls: From Village to City in a Changing China*. Random House Digital, Inc.

Choudhury, P. and T. Khanna (2014). Firms and Within Country Migration: Evidence from India.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications. *Journal of Economic Literature 51*(1), 5–62.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics 128*(2), 531–580.

Crépon, B. and G. J. Van den Berg (2016). Active Labor Market Policies. *Annual Review of Economics 8*, 521–546.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

Deaton, A. (1997). *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*. World Bank Publications.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature 47*(2), 315–372.

DellaVigna, S. (2018). Structural Behavioral Economics. *NBER Working Paper no. 24797*.

DellaVigna, S. and M. Gentzkow (2017). Uniform Pricing in US Retail Chains. *NBER Working Paper No. 23996*.

DellaVigna, S. and D. Pope (2016). What Motivates Effort? Evidence and Expert Forecasts. *NBER Working Paper No. 22193*.

Deserranno, E. (2019). Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda. *American Economic Journal: Applied Economics 11*(1), 277–317.

Diamond, A. (2013). Executive Functions. *Annual review of psychology 64*, 135–168.

Dohmen, T. J., A. Falk, D. Huffman, and U. Sunde (2012). Interpreting Time Horizon Effects in Inter-Temporal Choice. *Working Paper*.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology 92*(6), 1087.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2016). Global Evidence on Economic Preferences. *Working Paper*.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Working Paper*.

Flory, J. A., A. Leibbrandt, and J. A. List (2014). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. *The Review of Economic Studies 82*(1), 122–155.

Franklin, S. (2017, 11). Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies. *The Economic Journal 128*(614), 2353–2379.

Gabaix, X. and D. Laibson (2005). Bounded Rationality and Directed Cognition. *Working Paper*.

Galenianos, M., P. Kircher, and G. Virág (2011). Market Power and Efficiency in a Search Model. *International Economic Review 52*(1), 85–103.

Gardner, D. G. and J. L. Pierce (2010). The Core Self-Evaluation scale: Further Construct Validation Evidence. *Educational and Psychological Measurement 70*(2), 291–304.

Gobillon, L., H. Selod, and Y. Zenou (2007). The Mechanisms of Spatial Mismatch. *Urban studies 44*(12), 2401–2427.

Hardy, M. and J. McCasland (2017). Are Small Firms Labor Constrained? Experimental Evidence from Ghana. *Working Paper*.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013). The Allocation of Talent and US Economic Growth. *NBER Working Paper No. 18693*.

Hsieh, C.-T. and E. Moretti (2015). Why do Cities Matter? Local Growth and Aggregate Growth. *NBER Working Paper No. 21154*.

Imbert, C. and J. Papp (2016). Short-term Migration Costs: Evidence from India's Employment Guarantee. *Working Paper*.

John, O. P. and S. Srivastava (1999). The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. *Handbook of Personality: Theory and Research 2*(1999), 102–138.

Kaboski, J. P. and R. M. Townsend (2011). A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative. *Econometrica 79*(5), 1357–1406.

Kessler, J. B., C. Low, and C. D. Sullivan (2019). Incentivized Resume Rating: Eliciting Employer Preferences without Deception. *American Economic Review 109*(11), 3713–44.

Koenker, R. and K. Hallock (2001). Quantile Regression: An Introduction. *Journal of Economic Perspectives 15*(4), 43–56.

Kremer, M., G. Rao, and F. Schilbach (2018). Behavioral Development Economics. *Working paper*.

Laajaj, R. and K. Macours (2017). Measuring Skills in Developing Countries. *Working Paper*.

Lagakos, D., M. Mobarak, and M. E. Waugh (2017). The Welfare Effects of Encouraging Rural-Urban Migration. *Working Paper*.

Lefcourt, H. M. (1991). *Locus of Control*. Academic Press.

Machado, J. A. and J. S. Silva (2000). Glejser's Test Revisited. *Journal of Econometrics 97*(1), 189–202.

Machado, J. A. F., P. Parente, and J. M. C. Santos Silva (2011). QREG2: Stata Module to Perform Quantile Regression with Robust and Clustered Standard Errors.

Malmendier, U. and G. Tate (2015). Behavioral CEOs: The Role of Managerial Overconfidence. *The Journal of Economic Perspectives 29*(4), 37–60.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty Impedes Cognitive Function. *Science 341*(6149), 976–980.

Marimon, R. and F. Zilibotti (1999). Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. *The Economic Journal 109*(455), 266–291.

Mas, A. and A. Pallais (2016). Valuing Alternative Work Arrangements. *NBER Working Paper No. 22708*.

Mas, A. and A. Pallais (2017). Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field. *Working Paper*.

McKenzie, D. (2017a). How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. *The World Bank Research Observer*.

McKenzie, D. (2017b). Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review 107*(8), 2278–2307.

Mullainathan, S. and E. Shafir (2013). *Scarcity: Why Having Too Little Means So Much*. Macmillan.

Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *The American Economic Review 85*(5), 1313–1326.

Nichols, A. L. and R. J. Zeckhauser (1982). Targeting Transfers Through Restrictions on Recipients. *The American Economic Review 72*(2), 372–377.

Paserman, M. D. (2008). Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation. *The Economic Journal 118*(531), 1418–1452.

Phillips, D. C. (2014). Getting to work: Experimental evidence on job search and transportation costs. *Labour Economics 29*, 72–82.

Porzio, T. (2017). Cross-Country Differences in the Optimal Allocation of Talent and Technology.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive psychology 41*(1), 1–48.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-theoretic models of the labor market: A survey. *Journal of economic literature 43*(4), 959–988.

Rosenberg, M. (1986). *Conceiving the Self*. RE Krieger.

Saleheen, J., I. Levina, M. Melolinna, and S. Tatomir (2017). The Financial System and Productive Investment: New Survey Evidence. *Bank of England Quarterly Bulletin Q1*.

Santangelo, V. and E. Macaluso (2013). Visual Salience Improves Spatial Working Memory via Enhanced Parieto-Temporal Functional Connectivity. *Journal of Neuroscience 33*(9), 4110–4117.

Schmidt, F. L. and J. E. Hunter (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological bulletin 124*(2), 262.

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

Vansteelandt, S. (2009). Estimating Direct Effects in Cohort and Case–Control Studies. *Epidemiology 20*(6), 851–860.

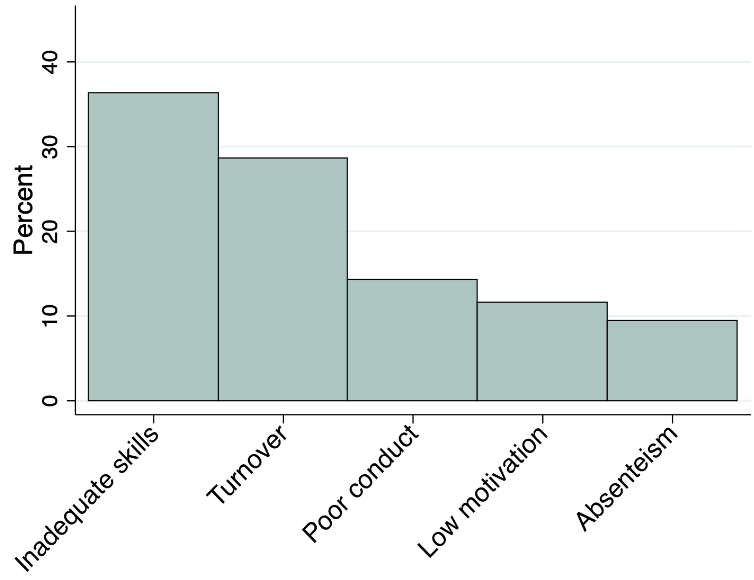
Weaver, J. (2016). Jobs for Sale: Bribery and Misallocation in Hiring. *Working Paper*.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Zenou, Y. (2009). *Urban Labor Economics*. Cambridge University Press.

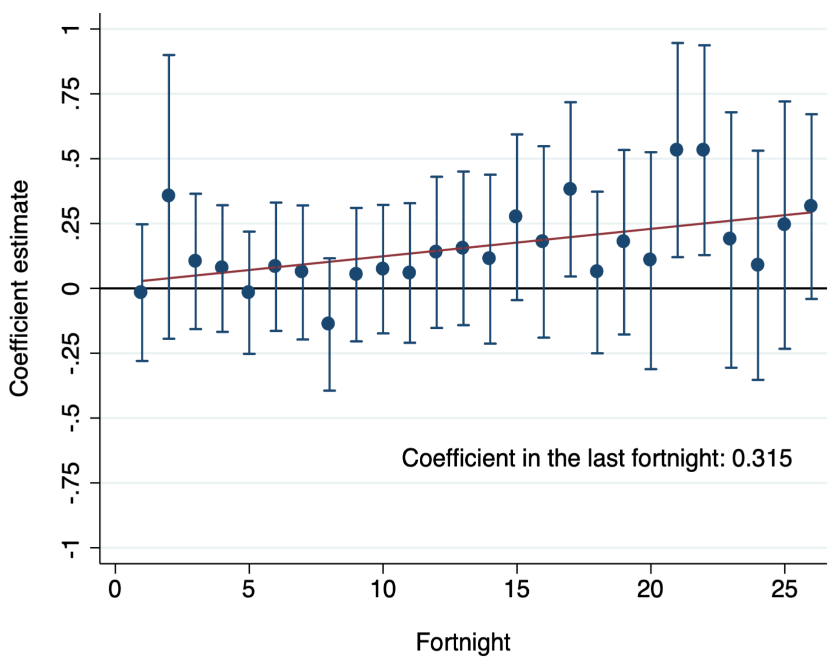
Figures and tables for inclusion in the main text

Figure 1: Most important HR problem



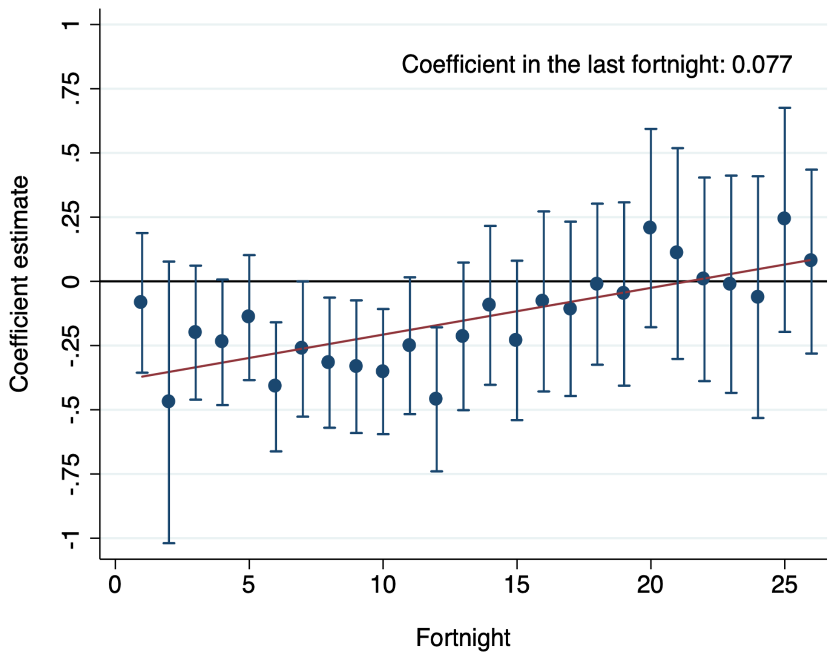
Notes: This figure reports data from our survey of firms hiring clerical workers. We report the distribution of managers' responses to the question 'What is the most important HR problem faced by your firm?'. Sample used: all managers.

Figure 2: Low savings and Raven test score among jobseekers
by fortnight



Notes: In this figure, we plot the coefficients from a battery of regressions of standardised Raven test scores on a dummy for having below-median savings. We estimate a separate regression for each fortnight, each time restricting the sample to those individuals who are out of work and searching for employment in that particular fortnight. This regression thus estimates the correlation between savings and Raven test scores among the individuals that are actively looking for employment in a given fortnight. Changes in this correlation are driven by selection in and out of job search. The data comes from the high-frequency panel of Abebe et al. (2020). The jobseekers in this dataset are interviewed on the phone every two weeks for a period of one year. The same jobseekers also complete face-to-face baseline and endline interviews before and after the phone survey. For logistical reasons, the first three fortnights and the last fortnight of the phone survey have much fewer observations than the other fortnights and we thus drop them. We include face-to-face baseline and endline observations and consider them as the first and last fortnight of the panel. For each regression, we report the point estimate of the coefficient on the dummy variable for having below-median savings and a 90 percent confidence interval obtained using robust standard errors.

Figure 3: High distance and Raven test score among jobseekers by fortnight



Notes: In this figure, we plot the coefficients from a battery of regressions of standardised Raven test scores on a dummy for living at a distance from the city centre above the median in the sample. We estimate a separate regression for each fortnight, each time restricting the sample to those individuals who are out of work and searching for employment in that particular fortnight. This regression thus estimates the correlation between distance and Raven test scores among the individuals that are actively looking for employment in a given fortnight. Changes in this correlation are driven by selection in and out of job search. The data comes from the high-frequency panel of Abebe et al. (2020). The jobseekers in this dataset are interviewed on the phone every two weeks for a period of one year. The same jobseekers also complete face-to-face baseline and endline interviews before and after the phone survey. For logistical reasons, the first three fortnights and the last fortnight of the phone survey have much fewer observations than the other fortnights and we thus drop them. We include face-to-face baseline and endline observations and consider them as the first and last fortnight of the panel. For comparability with Figure 2, we restrict the sample to observations with non-missing savings. Results are qualitatively unchanged if we do not apply this restriction. For each regression, we report the point estimate of the coefficient on the dummy variable for living at above-median distance from the city centre and a 90 percent confidence interval obtained using robust standard errors.

Figure 4: Illustration of the effect of an application incentive (noisy ability case, $\rho_z > 0$)

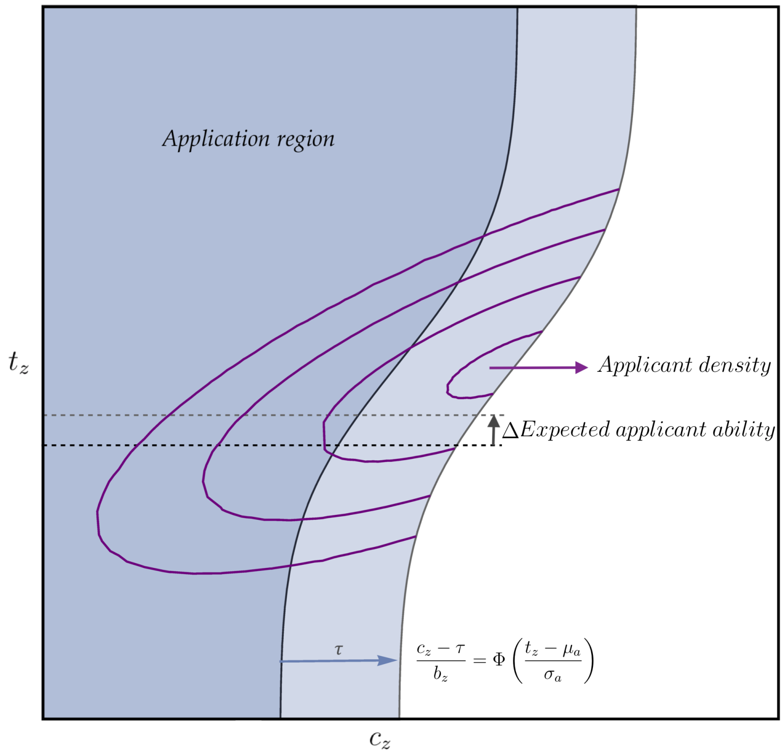


Figure 5: Illustration of the effect of an application incentive
(noisy ability case, $\rho_z > 0$)

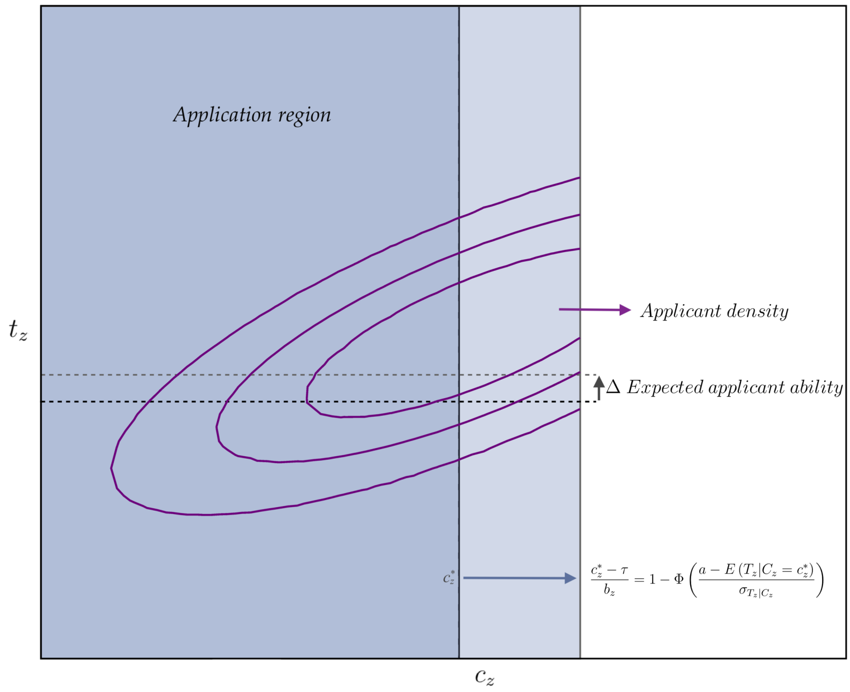
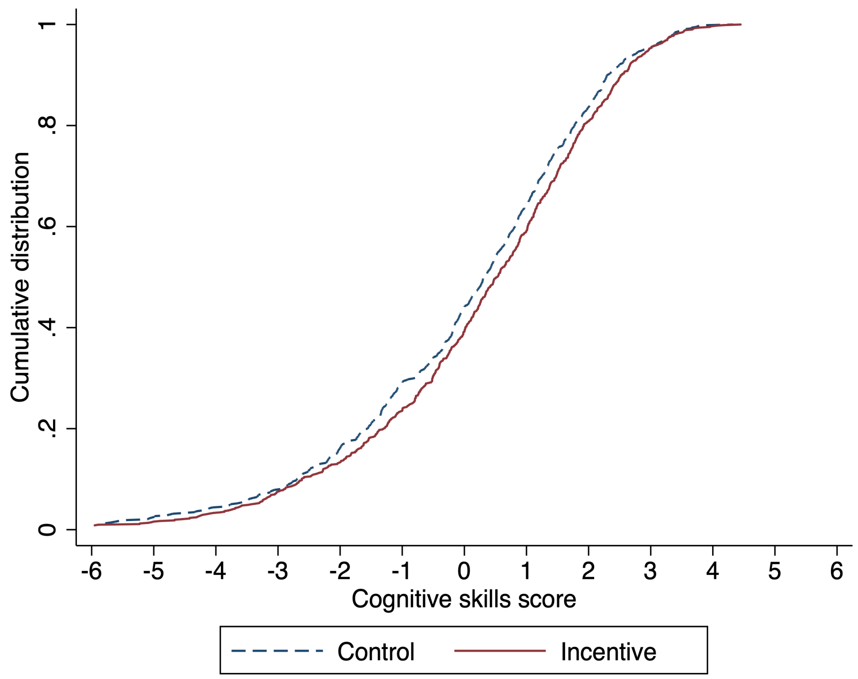
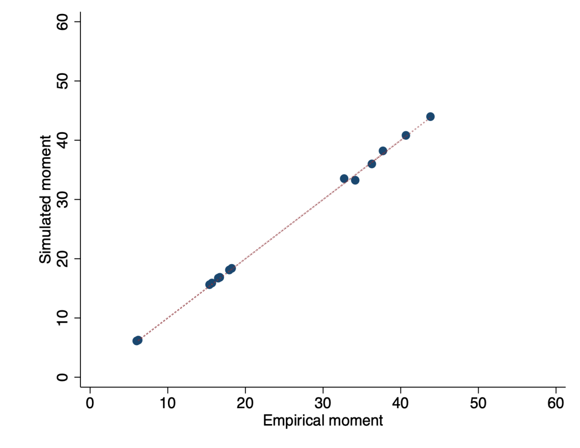


Figure 6: Impacts on the distribution of applicant cognitive ability
Incentive treatment

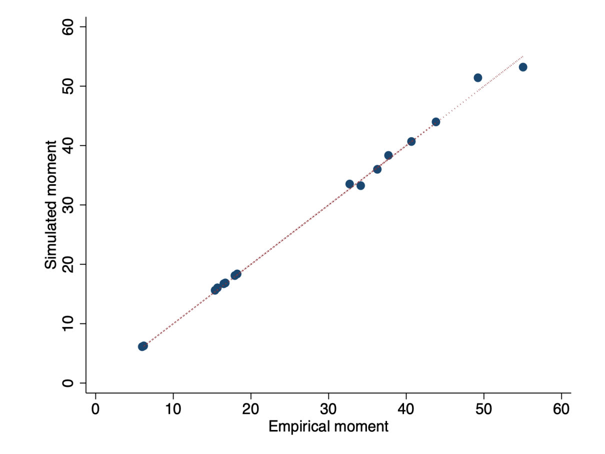


Notes: This figure plots the distribution of cognitive ability among control and application incentive applicants in the experiment. Sample used: all applicants (incentive and control groups).

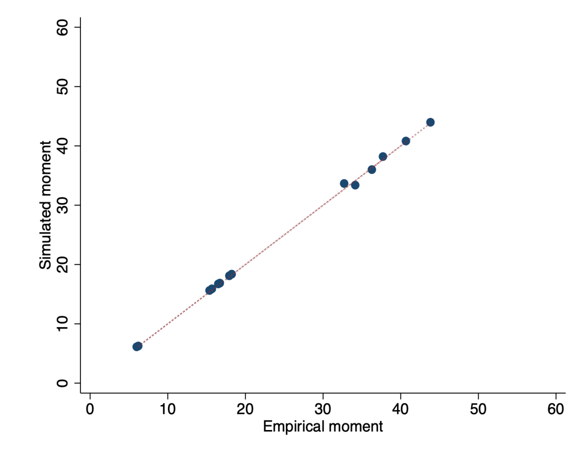
Figure 7: Moment fit



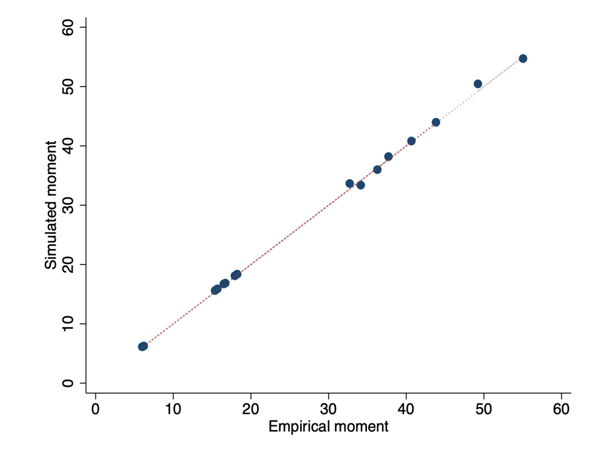
(a) Noisy ability, core moments



(b) Noisy ability, additional moments



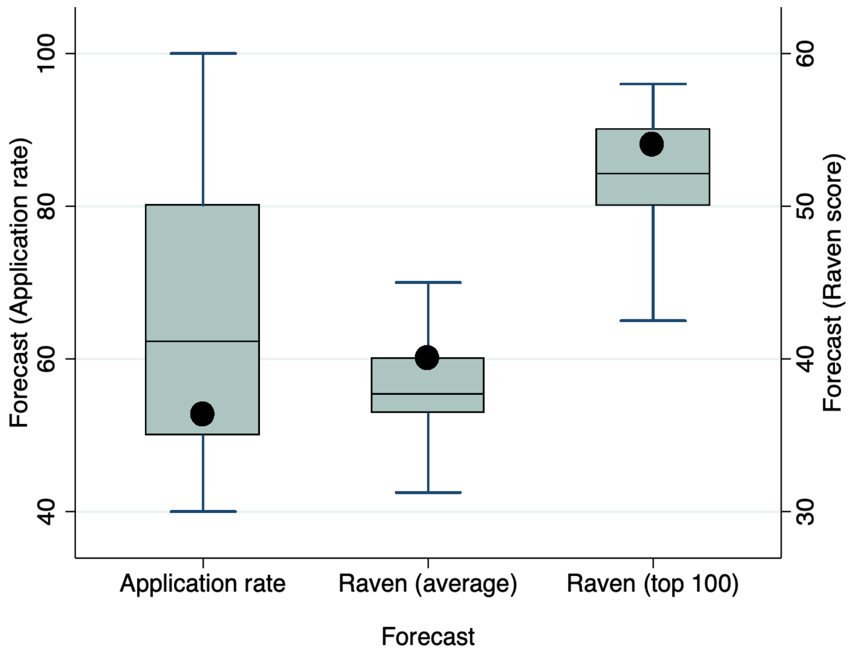
(c) Noisy selection, core moments



(d) Noisy selection, additional moments

Notes: This figure plots the simulated moments against the empirical moments for all models presented in Table 4. All moments are normalised using the standard deviation of the respective empirical moment (which we obtain through a bootstrap). The underlying moments are reported in Table A.49 -Table A.52 of the online Appendix. For ease of representation, we plot the application rate in each treatment group (as opposed to the treatment effect on application rates) and average ability in each treatment group (as opposed to the treatment effect on average ability).

Figure 8: Forecast accuracy of firm managers



Notes: This figure reports managers' forecasts about the outcomes of the experiment. The circle shows the true value of the variable forecasted. The box-plot shows the distribution of the forecasts: (i) the horizontal line inside the box shows the mean forecast, (ii) the box shows the interquartile range of the forecasts, (iii) the upper and lower caps on the whiskers show the minimum and maximum forecasts. Sample used: all managers.

Table 1: Summary statistics and balance

	Mean			StDev	N	Balance tests (p)		
	Control	Incentive	High wage	Control		Incentive = Wage = Control	Incentive = Control	Wage = Control
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.21	0.21	0.21	0.41	4689	0.98	0.88	0.88
Age	26.08	25.95	26.24	4.28	4686	0.21	0.41	0.32
Born in Addis Ababa	0.24	0.24	0.23	0.43	4689	0.53	0.72	0.45
First language is Amharic	0.68	0.70	0.67	0.47	4689	0.21	0.15	0.89
Heard about job in newspaper	0.55	0.58	0.56	0.50	4689	0.33	0.18	0.94
Engineering or hard science	0.50	0.49	0.48	0.50	4689	0.46	0.35	0.24
Economics	0.15	0.16	0.17	0.36	4689	0.53	0.38	0.29
Other social science	0.15	0.16	0.14	0.36	4689	0.15	0.50	0.22
Wage work experience (dummy)	0.53	0.53	0.53	0.50	4689	0.97	0.89	0.82
Wage work experience (months)	28.12	28.45	29.06	43.53	4689	0.84	0.84	0.56
Self-employed experience	0.33	0.35	0.35	0.47	4689	0.59	0.38	0.37
Currently unemployed	0.67	0.65	0.64	0.47	4689	0.18	0.22	0.07
Currently wage employed	0.24	0.26	0.27	0.43	4689	0.18	0.28	0.07
GPA	2.97	2.99	2.96	0.46	4451	0.22	0.18	0.79
Overall balance						0.33	0.61	0.47

Notes: Balance tests. In columns 1-5, we report descriptive statistics and sample sizes. In columns 6-8, we report the p-values of a battery of balance tests. In column 6, we test the null hypothesis that the three experimental groups are balanced. In columns 7 and 8, we present pair-wise tests of a single treatment group against the control. In the last row, we report a joint test of orthogonality (following the recent literature, e.g. McKenzie (2017b)). To perform the joint test of orthogonality we regress the treatment variable on all covariates and we then test the joint null hypothesis that all covariates have a zero coefficient. We do this by estimating a categorical logit model in column 6 (the treatment variable can take three values) and OLS models in columns 7 and 8 (here we drop one experimental group, so the treatment variable only takes two values). Sample used: baseline sample.

Table 2: Application rates

	(1)
Incentive	0.115
	(0.016)
High Wage	0.187
	(0.016)
Control mean	0.412
Incentive = Wage (p)	0.000
Obs.	4689

Notes: OLS regression. The dependent variable is a dummy capturing whether the respondent has applied to the experiment's job. The second to last row reports the p-value of a test of the null hypothesis that the two treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: baseline sample.

Table 3: Cognitive ability

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.248	.229	.229	.17	.412	.079
	(0.112)	(0.110)	(0.117)	(0.133)	(0.173)	(0.250)
	[0.081]	[0.115]	[0.148]	[0.607]	[0.053]	[1.000]
High Wage	.194	.202	.227	.075	.28	.155
	(0.110)	(0.108)	(0.112)	(0.131)	(0.165)	(0.227)
	[0.225]	[0.182]	[0.130]	[0.852]	[0.271]	[0.743]
Control value	-0.0000	2.312	1.477	0.356	-1.238	-2.697
Incentive = Wage (p)	0.574	0.795	0.983	0.448	0.371	0.741
Obs.	2386	2386	2386	2386	2386	2386

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The dependent variable is the index of cognitive ability. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) are reported in brackets. q-values control the false discovery rate for the multiple tests of the same hypothesis for different indices of ability. A Wilcoxon rank-sum test rejects the equality of the distribution of cognitive ability in the control and incentive groups (p=.038) and marginally fails to reject the equality of the distribution of cognitive ability in the control and wage groups (p=.107). Sample used: all applicants.

Table 4: Parameter estimates

	(1)	(2)	(3)	(4)
	Low B			
μ_C	136.320	187.950	272.950	206.800
	(20.69)	(15.99)	(32.21)	(234.69)
σ_C	192.840	192.010	200.480	208.570
	(35.90)	(42.51)	(317.59)	(1839.90)
ρ	0.640	0.643	0.612	0.625
	(0.06)	(0.05)	(0.21)	(0.22)
	High B			
μ_C	217.310	287.910	383.360	286.170
	(15.66)	(14.53)	(52.01)	(60.86)
σ_C	245.100	239.910	256.060	263.320
	(38.40)	(49.90)	(287.95)	(1182.06)
ρ	0.572	0.581	0.629	0.608
	(0.05)	(0.05)	(0.23)	(0.25)
a	50.428	46.934		
	(1.19)	(0.82)		
μ_a			-4.2e+03	792.460
			(2.19)	(9.07)
σ_a			1.2e+04	1.4e+04
			(0.86)	(0.69)
τ	19.063	15.781	56.556	58.556
	(15.93)	(13.80)	(91.39)	(358.42)
τ^w	55.732	35.763	110.350	153.570
	(53.50)	(87.08)	(306.60)	(1170.18)
Information	Noisy ability		Noisy selectivity	
Moments	14	16	14	16
Goodness of fit	2.2637	2.3389	2.2629	2.2654

Notes: The table shows parameter estimates obtained using minimum distance estimation. Columns (1) and (2) report estimates for the noisy-ability case. Columns (3) and (4) report estimates for the noisy-selectivity case. The estimates reported in columns (1) and (3) use fourteen empirical moments. The estimates reported in columns (2) and (4) use sixteen empirical moments. All empirical and simulated moments are reported in Tables A.49 - A.52. The parameters that describe the marginal distribution of ability are omitted for brevity and are reported in Table A.48 in the online Appendix. Standard errors obtained through a bootstrap of the structural estimation reported in parenthesis. The bootstrap includes the estimation of B and the demediation procedure. Costs are expressed in Ethiopian Birr.

Table 5: The returns of the interventions

	Incentive given to...				High wage
	All applicants	All hires	Female	Young	
	(1)	(2)	(3)	(4)	(5)
Internal Rate of Return	9.8	382.5	86.5	47.1	<0
	(-18.7, 38.3)	(44.6, 720.4)	(4.3, 168.7)	(24.6, 69.6)	
			Costs		
Time costs (month 0)	386	133	77	207.9	485
Incentive (month 0)	5862	300	1017	2439.7	0
Wage costs (months 1-3)	0	0	0	0	4800
			Benefits		
Value of higher ability (months 1-45)	168	61	65	114.2	211

Notes: Cost-benefit analysis. We consider five interventions: (i) the application incentive as implemented in the experiment (column 1), (ii) an incentive offered conditional on meeting the selectivity threshold set by the employer in the application test (column 2), (iii) an incentive offered to all women (column 3), (iv) an incentive offered to all individuals below the median age (column 4), (v) the high wage intervention as implemented in the experiment (column 5). The figures reported in columns 1, 2 and 5 are based on the parameters estimated for the noisy-ability case, using the core set of moments. The first row reports the Internal Rate of Return (IRR) of each intervention. Below the IRR we report a 90 percent confidence interval obtained with a bootstrap. In the remaining rows we report costs and benefits expressed in Ethiopian Birr. The benefit of the intervention is given by the increase in the expected cognitive ability of hires, multiplied by the return to cognitive ability estimated using local labour market data. We consider three types of potential costs: (i) the time required to assess the additional applications, (ii) the cost of the incentive, (iii) the wage increase. All assumptions used in the computation of costs and benefits are discussed in detail in Appendix A.4.

Table 6: Firm managers' ranking of workers

	Ranked first		Ranked first or second
	(1)	(2)	(3)
Incentive	0.369	0.456	0.374
	(0.053)	(0.064)	(0.055)
High wage	0.154		0.287
	(0.048)		(0.062)
Control mean	0.159	0.272	0.446
Incentive = Wage (p)	0.001		0.069
No. managers	195	195	195
Obs.	585	390	585

Notes: OLS regression. The dependent variable is indicate in the column's heading. The unit of observation is an applicant-manager pair. We thus have three observations per manager. In column (2) we drop applicants from the high wage group. Standard errors clustered at the manager level reported in parenthesis. Sample used: all managers.

Appendix
(For Online Publication)

A.1 Proofs

Proposition 2 *Suppose (T, B, C) are observable and distributed according to Assumptions 1 and 2. Further assume jobseekers anticipate that the threshold necessary to get the job is $a \sim \mathcal{N}(\mu_a, \sigma_a)$. Then it follows that for each $B = b_z > 0$, the application incentive (i) increases application rates, and (ii) increases the average ability of applicants, whenever*

$$\frac{\sigma_{T_z}}{\sigma_a \sigma_{C_z}} \frac{b_z}{\sqrt{2\pi}} \leq \rho_z$$

Proof. The application incentive is modeled as a shock that lowers application costs, shifting the distribution of C_z by an amount τ for all B-types; so the proof requires exploring the conditions under which application rates and expected applicant ability are increasing in τ. Under Assumptions 1 and 2, this exploration is identical for each of the B-types, so we drop here the subscripts without loss of generality.

(i) **Increase in application rates.** Using $\mathbb{I}_E$ to denote the indicator function for the application event, we can write the application rate as

$$\begin{aligned}
&\Pr\left(C \leq \Phi\left(\frac{T-\mu_a}{\sigma_a}\right) b + \tau\right) \\
&= \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}_E \cdot f(c|t) f(t)\, dc\, dt\right) \\
&= \left(\int_{-\infty}^{\infty} \Phi\left(\frac{\Phi\left(\frac{t-\mu_a}{\sigma_a}\right) b + \tau - E(C|T=t)}{\sigma_{C|T}}\right) f(t)\, dt\right)
\end{aligned}$$

Hence, differentiating with respect to τ gives

$$\frac{1}{\sigma_{C|T}} \left(\int_{-\infty}^{\infty} \phi\left(\frac{\Phi\left(\frac{t-\mu_a}{\sigma_a}\right) b + \tau - E(C|T=t)}{\sigma_{C|T}}\right) f(t)\, dt\right)$$

which is strictly positive.

(ii) **Increase in the average ability of applicants.** Using once again the same notation, we can write the expected ability of applicants as

$$\begin{aligned}
&E\left(T \mid C \leq \Phi\left(\frac{T-\mu_a}{\sigma_a}\right) b + \tau\right) \\
&= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t \cdot \mathbb{I}_E \cdot f(c|t) f(t)\, dc\, dt}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{I}_E \cdot f(c|t) f(t)\, dc\, dt} \\
&= \frac{\int_{-\infty}^{\infty} t \cdot \Phi\left(Y(t;\tau)\right) f(t)\, dt}{\int_{-\infty}^{\infty} \Phi\left(Y(t;\tau)\right) f(t)\, dt}
\end{aligned}$$

where $Y(t;\tau) = \frac{\Phi\left(\frac{t-\mu_a}{\sigma_a}\right)b+\tau-E(C|T=t)}{\sigma_{C|T}}$

Hence, differentiating with respect to τ gives

$$K\left(\int_{-\infty}^{\infty} t\cdot\phi\left(Y(t;\tau)\right)f(t)\,dt - \frac{\int_{-\infty}^{\infty} t\cdot\Phi\left(Y(t;\tau)\right)f(t)\,dt}{\int_{-\infty}^{\infty}\Phi\left(Y(t;\tau)\right)f(t)\,dt}\int_{-\infty}^{\infty}\phi\left(Y(t;\tau)\right)f(t)\,dt\right)$$

where $K \equiv \frac{1}{\sigma_{C|T}}\left(\int_{-\infty}^{\infty}\Phi\left(Y(t;\tau)\right)f(t)\,dt\right)^{-1}$

Since the term K is strictly positive, the derivative of expected applicant ability with respect to τ will be positive whenever the ratio $\frac{\phi(Y(t;\tau))}{\Phi(Y(t;\tau))}$ is increasing in t.[50]

In general, the Inverse Mills Ratio $\frac{\phi(x)}{\Phi(x)}$ is decreasing in x; so a sufficient condition for a positive derivative is that $Y(t;\tau)$ is decreasing in t. That is

$$\frac{b}{\sigma_a}\phi\left(\frac{t-\mu_a}{\sigma_a}\right) \leq \frac{\partial E(C|T=t)}{\partial t}$$

Since $E(C|T=t) = \mu_C + \frac{\sigma_C}{\sigma_T}\rho(t-\mu_T)$ for the conditional bivariate normal distribution, the condition becomes

$$\frac{\sigma_T}{\sigma_C}\frac{b}{\sigma_a}\phi\left(\frac{t-\mu_a}{\sigma_a}\right) \leq \rho$$

and since the standard normal density $\phi(\cdot)$ has a maximum, this is achieved when

$$\frac{\sigma_T}{\sigma_a\sigma_C}\frac{b}{\sqrt{2\pi}} \leq \rho$$

$\square$

[50]Notice that whenever the ratio $\lambda = \frac{\phi(Y(t;\tau))}{\Phi(Y(t;\tau))}$ is increasing in t, then t and λ are positively correlated, so

$$\int_{-\infty}^{\infty} t\cdot\lambda\left(Y(t;\tau)\right)\left(\frac{\Phi\left(Y(t;\tau)\right)f(t)}{\int_{-\infty}^{\infty}\Phi\left(Y(t;\tau)\right)f(t)\,dt}\right)dt \geq$$
$$\int_{-\infty}^{\infty} t\left(\frac{\Phi\left(Y(t;\tau)\right)f(t)}{\int_{-\infty}^{\infty}\Phi\left(Y(t;\tau)\right)f(t)\,dt}\right)dt\int_{-\infty}^{\infty}\lambda\left(Y(t;\tau)\right)\left(\frac{\Phi\left(Y(t;\tau)\right)f(t)}{\int_{-\infty}^{\infty}\Phi\left(Y(t;\tau)\right)f(t)\,dt}\right)dt$$

Proposition 3 *Suppose (T, B, C) are distributed according to Assumptions 1 and 2. Assume jobseekers are confident about the selection threshold a, but they only observe C and B, which they can use to update their beliefs about the probability that they will pass the recruitment test $T > a$. Then for each $B = b_z > 0$, the application incentive (i) increases application rates, and (ii) increases the average ability of applicants, if and only if*

$$0 < \rho_z < \frac{\sqrt{2\pi}\sqrt{1-\rho_z^2}\sigma_{C_z}}{b_z}$$

Proof. The proof proceeds in 4 steps. In each step, the reasoning applies to all B-types; so we drop again the subscripts without loss of generality.

1. **Cut-off existence.** Let us define $H(c) = \Pr(T_z > a | C_z = c_z) - \frac{c_z}{b_z}$.

   Since $H(0 + \epsilon) > 0$ for some small positive ϵ, and $H(b - \epsilon') < 0$ for some positive ϵ', it must be the case that $H(c) = 0$ at least once as c traverses the interval $(0, b)$.

2. **Cut-off uniqueness.** Given cut-off existence, to show that the threshold c^* is unique it suffices to show that $H(c)$ is decreasing in c. Using a standard result from the conditional bivariate normal distribution, we have

$$H(c) = 1 - \Phi\left(\frac{a - \mu_T - \dfrac{\sigma_T}{\sigma_C}\rho(c - \mu_C)}{\sqrt{1-\rho^2}\sigma_T}\right) - \frac{c}{b}$$

   Hence, differentiating with respect to c gives

$$\frac{\rho}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_C}\exp\left\{-\frac{\left[a - \mu_T - \dfrac{\sigma_T}{\sigma_C}\rho(c - \mu_C)\right]^2}{2(1-\rho^2)\sigma_T^2}\right\} - \frac{1}{b}$$

   From this expression it is easy to check that:

   (a) when $\rho < 0$, the derivative is always negative so $H(c)$ has at least one root, which by monotonicity we know is unique;

   (b) when $\rho = 0$, $\alpha(c)$ is horizontal; so a similar argument applies, and the root is unique; and

   (c) when $\rho > 0$, the derivative is negative whenever $\rho < \frac{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_C}{b}$.

3. **Treatment effect on applications.** For the treatment group the threshold c^* is defined as the level of costs for which $H(c^*; \tau') = Pr(T_z > a | C_z = c_z^*) - \frac{c_z^* - \tau}{b_z} = 0$.

Hence, using implicit differentiation gives

$$\frac{dc^*}{d\tau'} = -\frac{\partial H(c^*;\tau')/\partial\tau'}{\partial H(c^*;\tau')/\partial c^*}$$

which is strictly positive.

Clearly, since the application threshold is increasing in τ, the share of applicants with costs lower than this threshold will also be increasing in τ.

4. **Treatment effect on the average ability of applicants.** Using the law of iterated expectations, we have that

$$\begin{aligned}
E(T|C<c^*) &= E(E(T|C)|C<c^*) \\
&= E\left(\mu_T + \frac{\sigma_T}{\sigma_C}\rho(C-\mu_C) \mid C<c^*\right) \\
&= \mu_T - \frac{\sigma_T}{\sigma_C}\rho\ \left(\mu_c - E\left(C \mid C<c^*\right)\right) \\
&= \mu_T - \rho\,\sigma_T\,\frac{\phi\left(\frac{c^*-\mu_C}{\sigma_C}\right)}{\Phi\left(\frac{c^*-\mu_C}{\sigma_C}\right)} \\
&= \mu_T - \rho\,\sigma_T\,\lambda\left(\frac{c^*-\mu_C}{\sigma_C}\right)
\end{aligned}$$

and differentiating with respect to τ gives

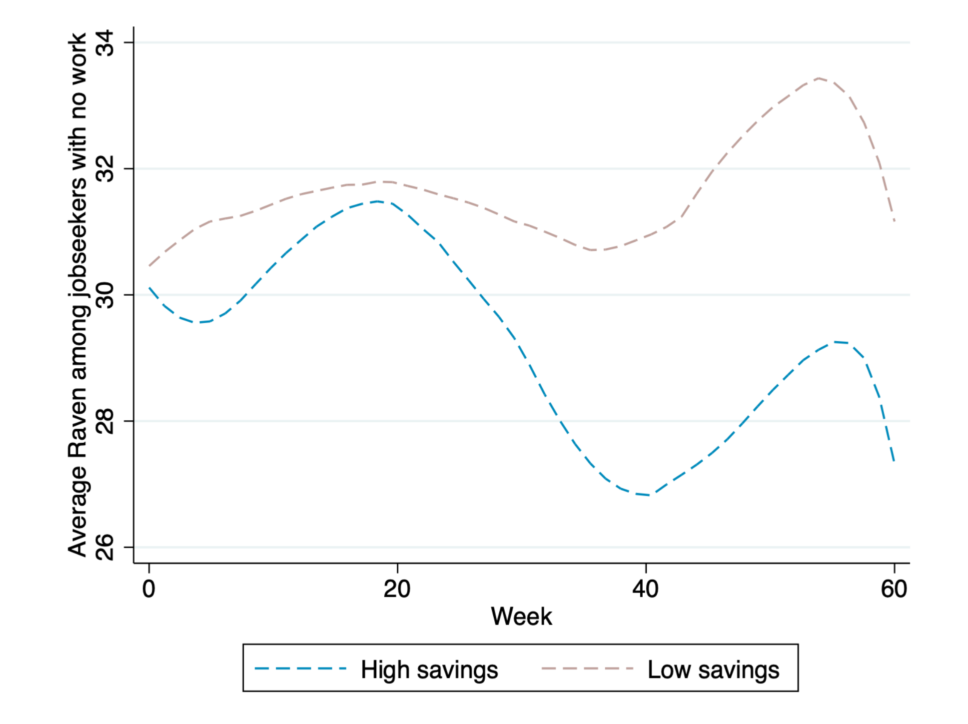
$$\frac{d}{d\tau}E(T|C<c^*(\tau)) = -\rho\frac{\sigma_T}{\sigma_C}\frac{dc^*}{d\tau}\frac{d\lambda(c)}{dc}$$

Since the Inverse Mills Ratio $\lambda(c)$ is decreasing in c, and we have shown that c^* is increasing in τ, the derivative is positive if and only if ρ is positive.

$\square$

A.2 Figures and tables

Figure A.1: The selection mechanism

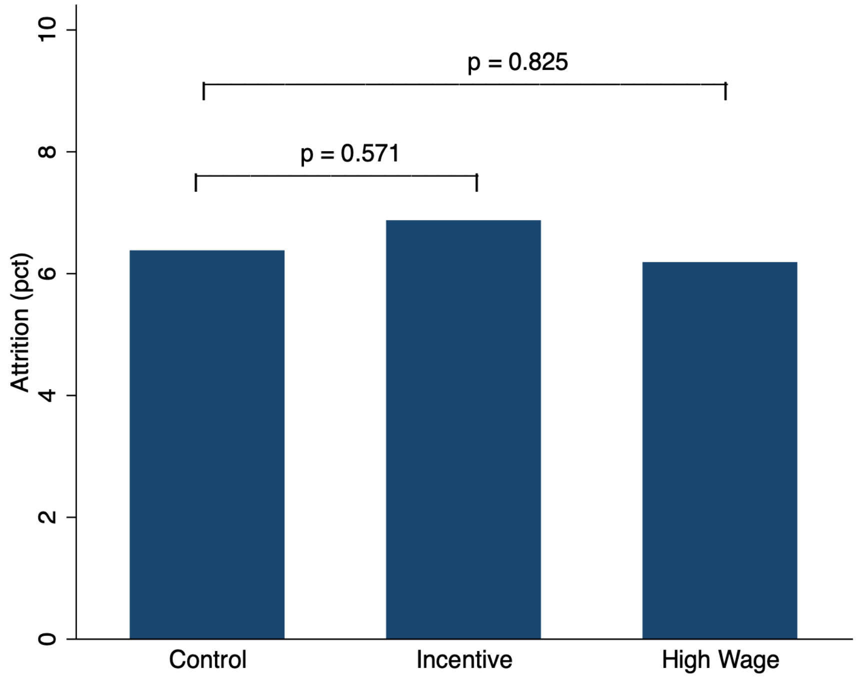


Notes: This figure reports data from the high-frequency survey of Abebe et al. (2020). We plot the average Raven test score among jobless individuals who are searching for work at a given point in time. Changes in this variable are due to movements in and out of job search over time. The figure is produced using Stata's `lpoly` command.

Figure A.2: Timeline

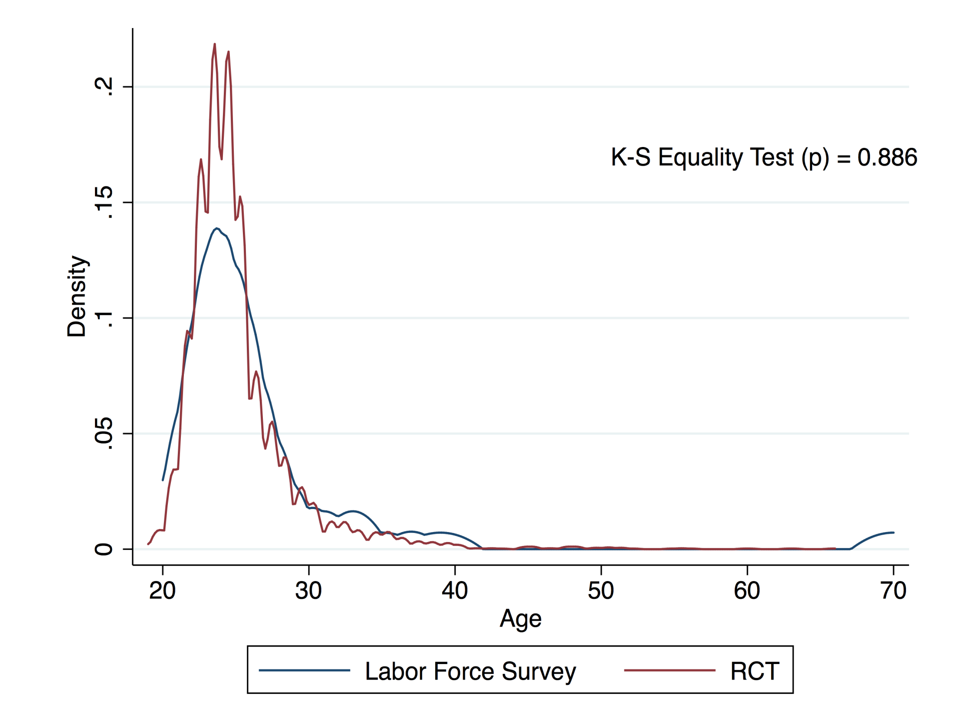
Notes: This figure shows the timeline of a typical hiring round. The position is advertised at the start of the week (day 1 in the timeline). Jobseekers can call to inquire about the position until close-of-business on Friday of that week (day 5). Those jobseekers who call to inquire about the position are then invited to make an in-person application on a randomly assigned day on the following week (days 7 to 12) or on the Monday of the third week (day 15). We thus have 6 application days, 2 of which are assigned to each experimental group. Finally, all jobseekers who call to inquire about the position are called again 30 days after the initial phone call. The jobseekers who are invited for an interview are told about this in a separate phone call shortly before the second phone call. Interviews are held shortly thereafter and the position starts right away. We collect data on jobseekers at each stage of the experiment. During the first phone call, we collect data on their socio-demographic characteristics, labour market experience and GPA. If jobseekers apply for the position, they have to complete several tests of ability (Raven and Stroop), and answer psychometric questions and questions about their economic preferences. Finally, in the second phone call, jobseekers are asked about job search in the last 30 days.

Figure A.3: Attrition



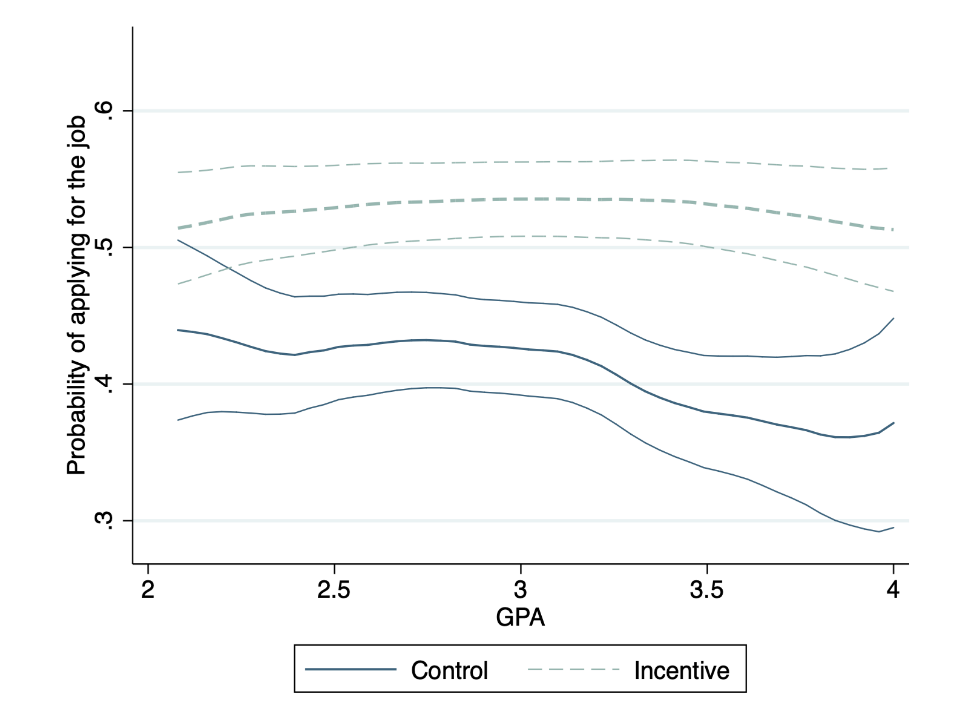
Notes: This figure shows descriptive data on attrition in the three experimental groups. An individual is considered attrited when our team is unable to contact them for the 30-days follow-up phone survey. The bars indicate total attrition in each group. Further, we report the p-values for a test of the null hypothesis of no differential attrition between a given treatment group and the control group. Sample used: baseline sample.

Figure A.4: Distribution of jobseeker age in experimental sample and in representative data



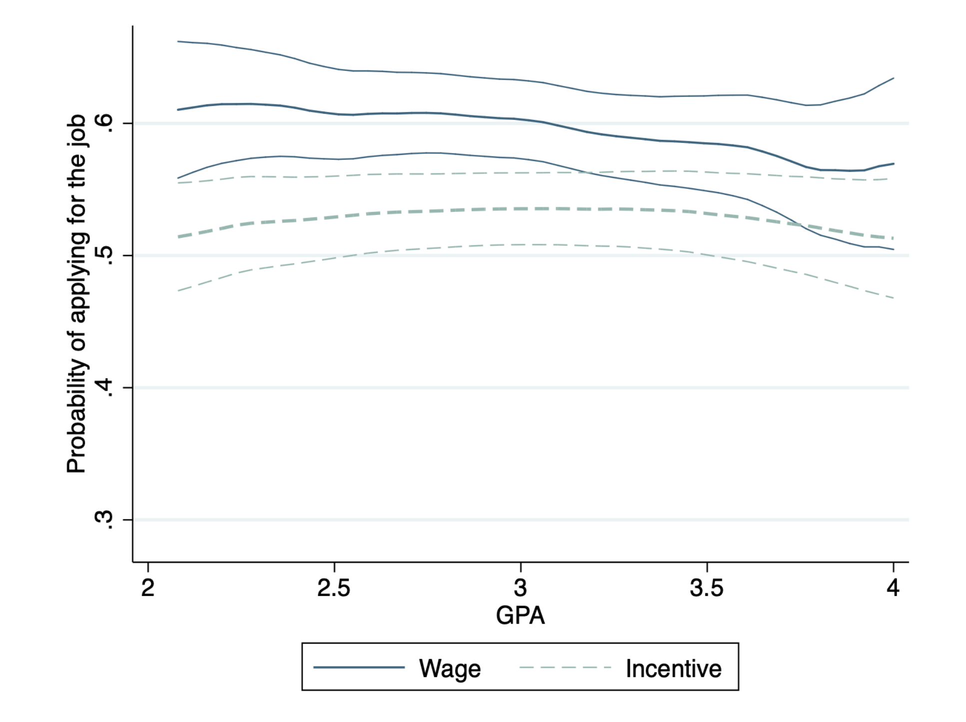
Notes: This figure shows kernel density plots of age in two samples (i) the experiment's sample and (ii) a representative sample of jobseekers in Addis Ababa (drawn from the 2013 Labour Force Survey of Ethiopia). The figure also reports the p-value of a Kolmogorov-Smirnov test of the equality of the two distributions. Samplea used: (i) baseline sample and (ii) sample from Ethiopia's 2013 Labour Force Survey.

Figure A.5: Applications and GPA score: Control and application incentive groups



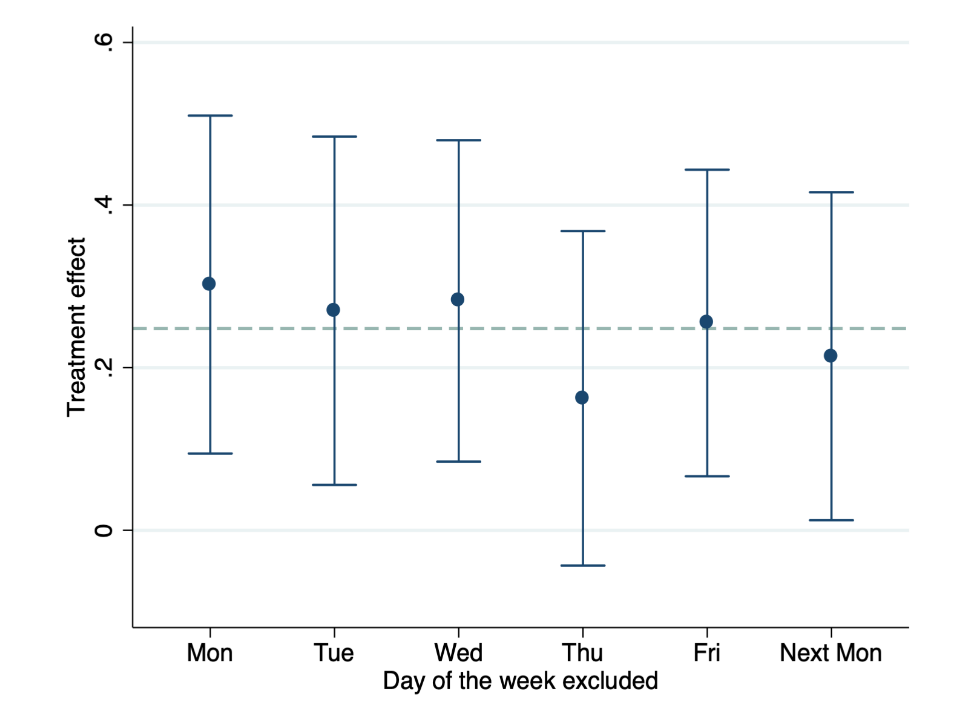
Notes: This figure shows the relationship between a jobseeker's GPA and the probability of applying for the experiment's job. We separately plot this relationship for individuals in the control and application incentive groups. We report 95 percent confidence intervals. The figure is produced using Stata's `lpolyci` command. Sample used: baseline sample (control and incentive groups).

Figure A.6: Applications and GPA score: Incentive and high wage groups



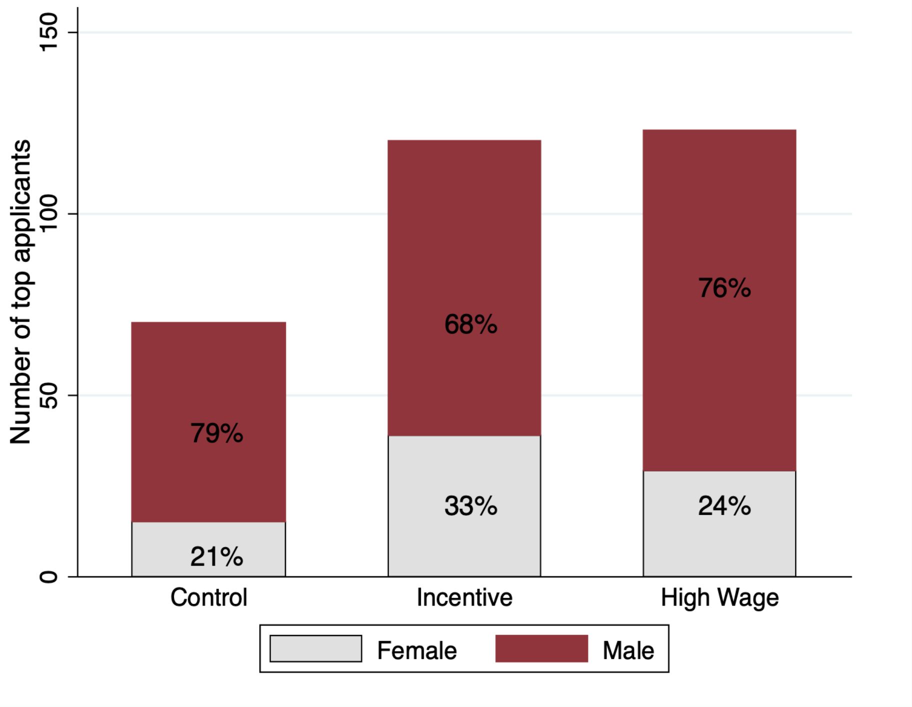
Notes: This figure shows the relationship between a jobseeker's GPA and the probability of applying for the experiment's job. We separately plot this relationship for individuals in the application incentive and high wage groups. We report 95 percent confidence intervals. The figure is produced using Stata's `lpolyci` command. Sample used: baseline sample (control and high wage groups).

Figure A.7: Robustness to exclusion of selected days of the week



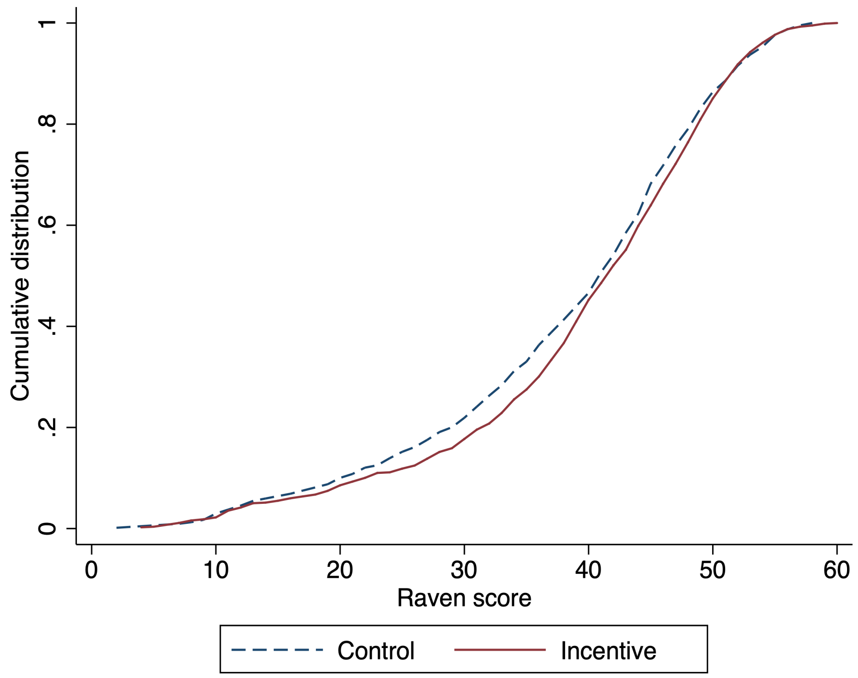
Notes: This figure shows the OLS estimates of the effect of the application incentive on average applicant ability for different samples. Each sample is obtained by dropping all individuals who are invited to take the test on a specific day of the week. The dashed horizontal line indicates the treatment effect for the full sample. Overall sample used: all applicants.

Figure A.8: The proportion of female top applicants



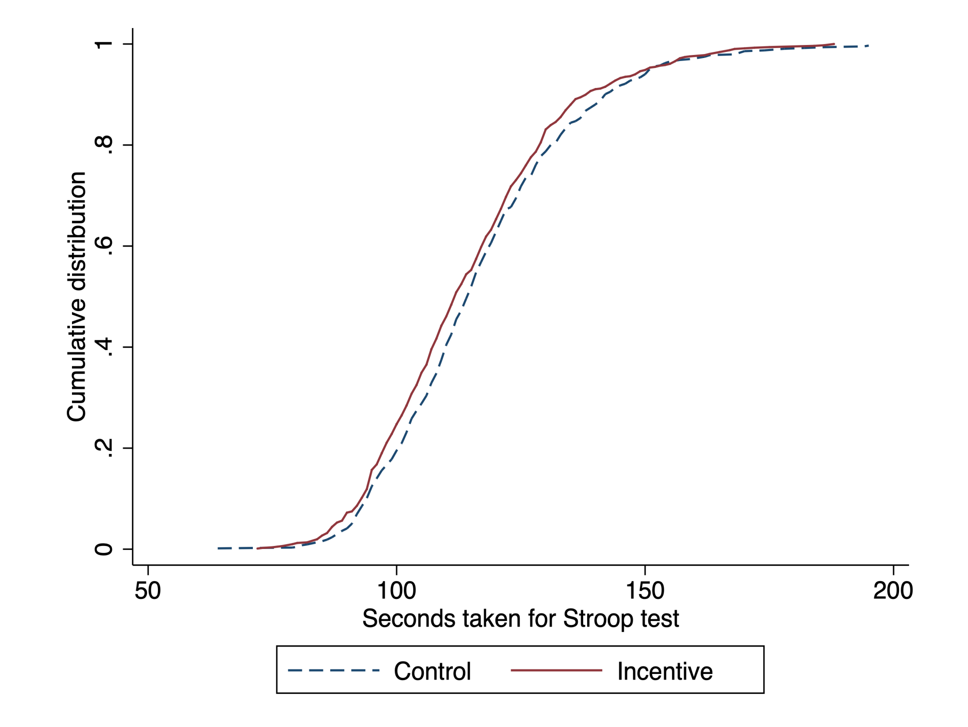
Notes: This figure plots the number of top applicants in each experimental group. Within each bar, we also report the proportion of top applicants in that experimental group that are female and the proportion of top applicants that are male. A 'top applicant' is defined as somebody whose cognitive ability is above the 90th percentile of the control group distribution of cognitive ability. Sample used: all top applicants.

Figure A.9: Impacts of incentives on the distribution of applicant Raven test score



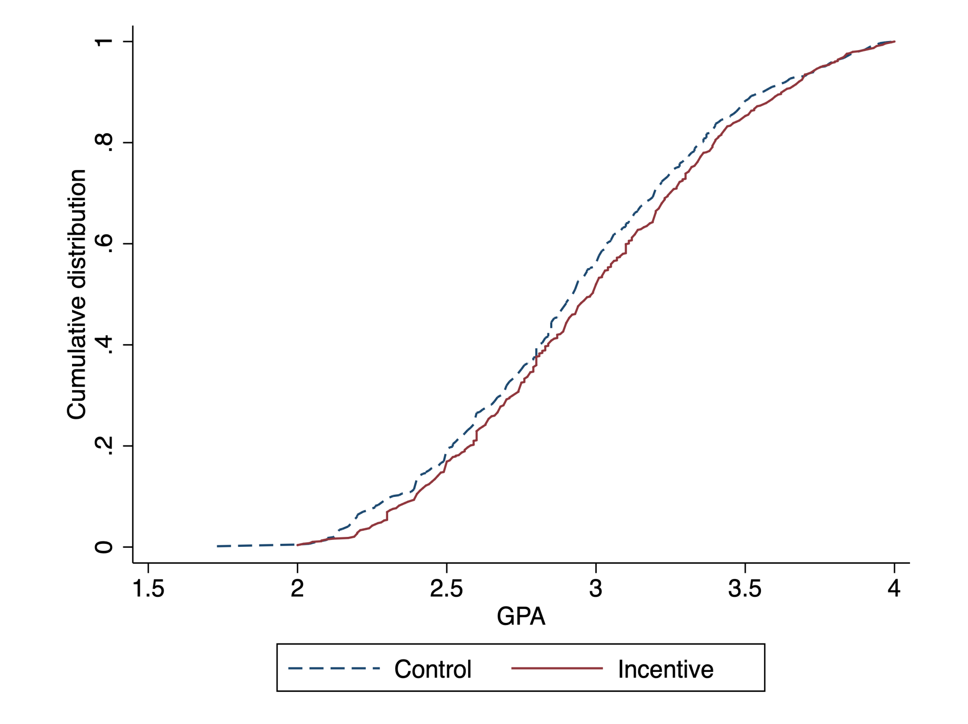
Notes: This figure plots the distribution of Raven test scores among control and application incentive applicants in the experiment. Sample used: all applicants (control and incentive groups).

Figure A.10: Impacts of incentives on the distribution of applicant Stroop test performance



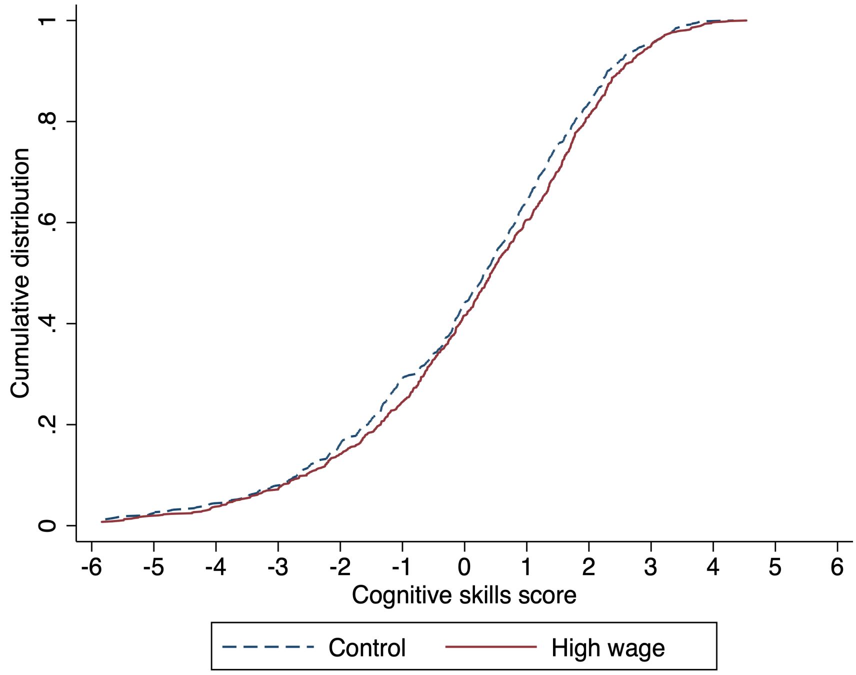
Notes: This figure plots the distribution of time taken to complete the Stroop test among control and application incentive applicants in the experiment. A smaller value indicates better performance. Sample used: all applicants (control and incentive groups).

Figure A.11: Impacts of incentives on the distribution of applicant GPA score



Notes: This figure plots the distribution of GPA scores among control and application incentive applicants in the experiment. Sample used: all applicants (control and incentive groups).

Figure A.12: Impacts of high wage offer on the distribution of applicant cognitive ability

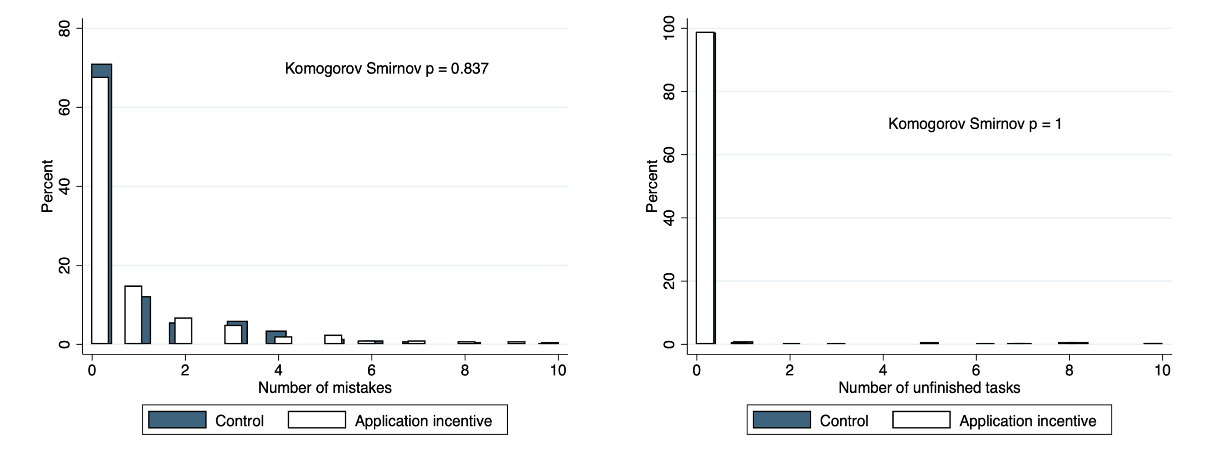


Notes: This figure plots the distribution of cognitive ability among control and application incentive applicants in the experiment. Sample used: all applicants (control and high wage groups).

Figure A.13: Distribution of test effort

(a) No. mistakes

(b) No. unfinished tasks

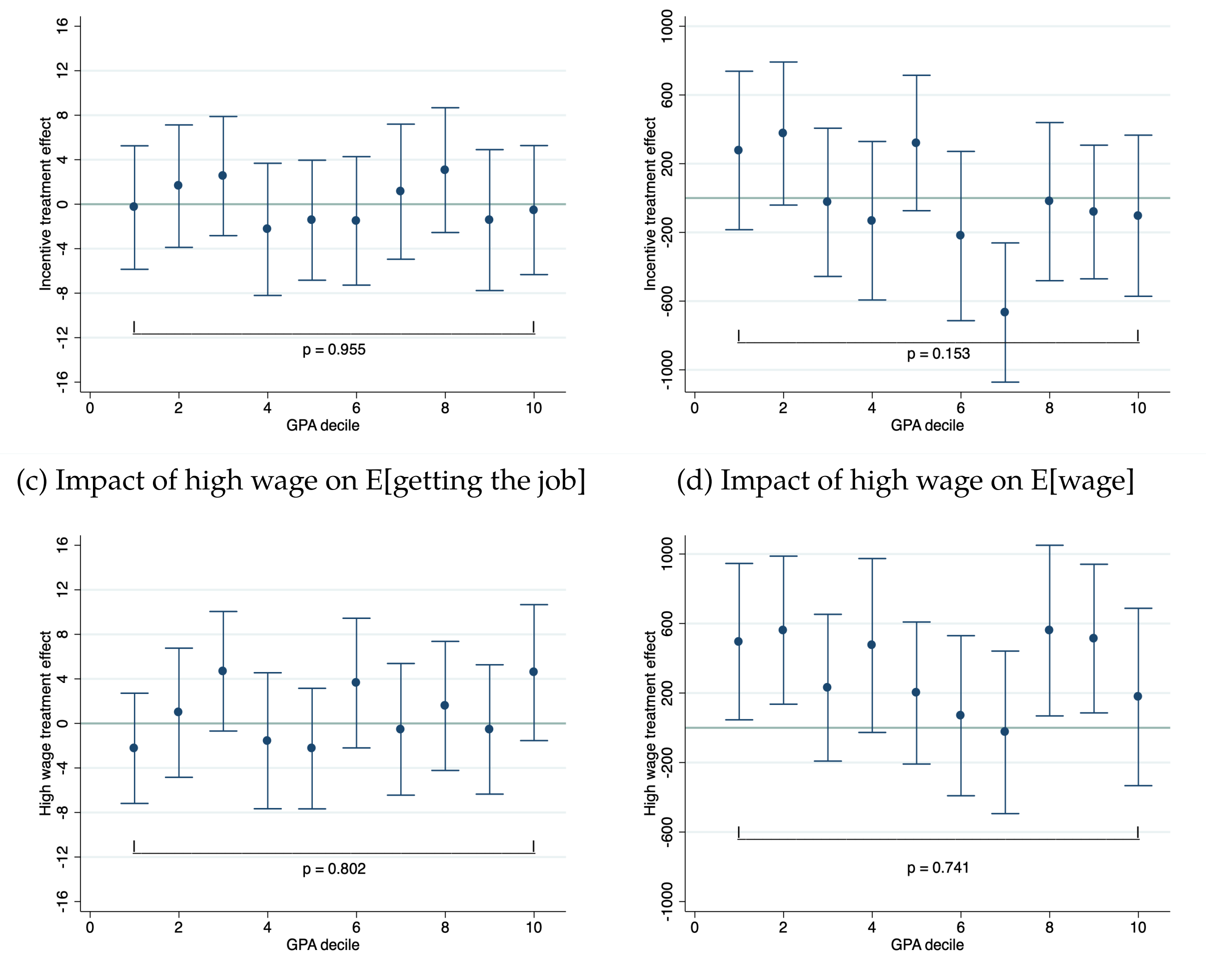


Notes: These figures show histograms of the two measures of test effort for applicants from the control and application incentive groups. The figures also report a p-value for a Kolmogorov Smirnov test of the equality of the two distributions. Sample used: all applicants (control and incentive groups).

Figure A.14: Belief updating and GPA

(a) Impact of incentive on E[getting the job] (b) Impact of incentive on E[wage]

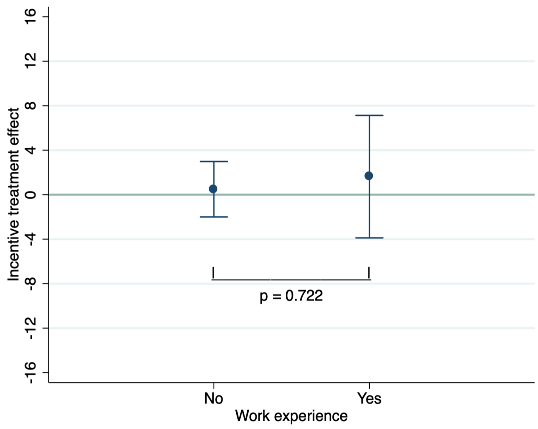
(c) Impact of high wage on E[getting the job] (d) Impact of high wage on E[wage]



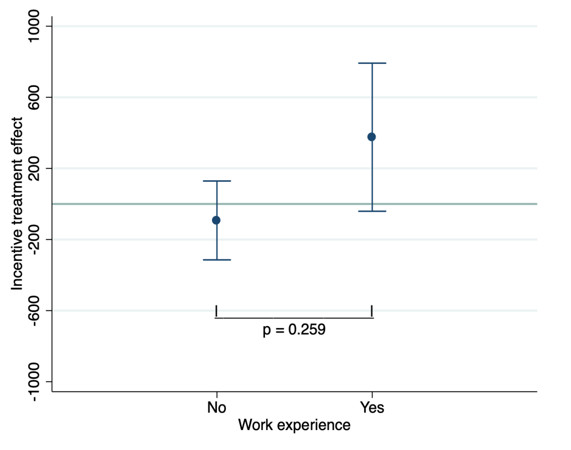
Notes: These figures show point estimates and 90% confidence intervals of the treatment effects on participants' beliefs about (i) the probability of getting the experiment's job, and (ii) the wage they would earn in their next job. Each figure also shows a p-value for a test of the null hypothesis that the coefficients are equal to each other. Sample used: baseline sample.

Figure A.15: Belief updating and experience

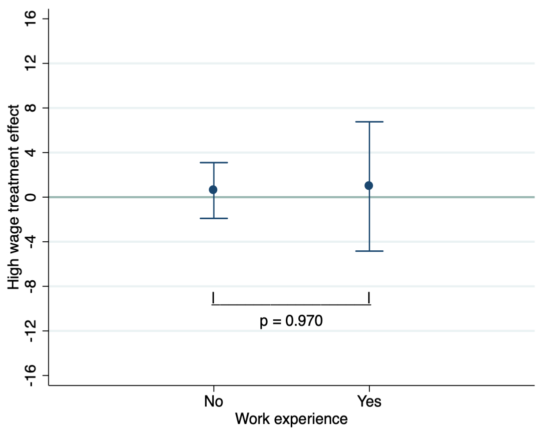
(a) Impact of incentive on E[getting the job]



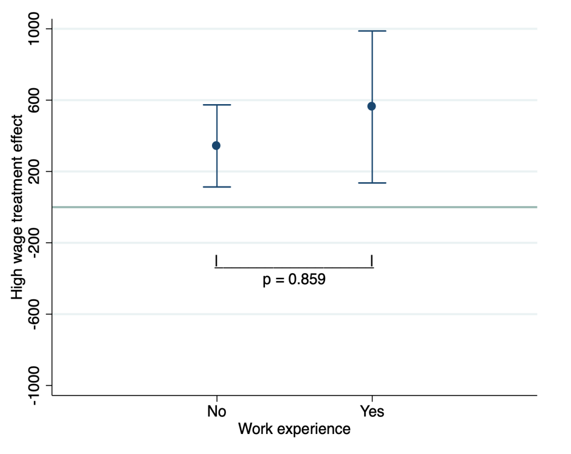
(b) Impact of incentive on E[wage]



(c) Impact of high wage on E[getting the job]



(d) Impact of high wage on E[wage]



Notes: These figures show point estimates and 90% confidence intervals of the treatment effects on participants' beliefs about (i) the probability of getting the experiment's job, and (ii) the wage they would earn in their next job. Each figure also shows a p-value for a test of the null hypothesis that the coefficients are equal to each other. Sample used: baseline sample.

Figure A.16: Sample CV

Candidate 1

Age: 33

Education

Highest level of education: BA (BSc) degree

Field of study: Natural and Computational Sciences

Average grade/GPA: 2.57

Work Experience

Has work experience? Yes

Last employer: MAYLEKO LOAGE

Type of employer: Private business

Test scores

Cognitive ability score: 410

Non-cognitive ability score: 600

Table A.1: Occupations included in the firm survey

Job code	Description	Examples
43-1	Supervisors of Support Workers	First-Line supervisors of office support workers
43-2	Communications Equipment Operators	Telephone operators
43-3	Financial Clerks	Bill and Account Collectors; Bookkeeping, accounting, and auditing clerks
43-4	Information and Record Clerks	Correspondence clerks, credit checkers, customer service representatives
43-9	Other Administrative Support Workers	Computer and data entry operators, claims processing clerks
13-1	Business Operations Specialists	Buyers and purchasing agents, cost estimators, claim checkers, logisticians
13-2	Legal Support Workers	Legal assistants, court workers

Table A.2: Ability and labour market outcomes

Dep. var.	Ln(wage)				Employed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven	.09			.09	.046			.046
	(0.035)			(0.035)	(0.018)			(0.018)
Conscientiousness	-.032	.019		.018		.007		.004
	(0.025)	(0.030)		(0.031)		(0.018)		(0.019)
Neuroticism	-.047		-.005	-.002			-.009	-.007
	(0.025)		(0.035)	(0.035)			(0.018)	(0.019)
Mean dep. var.	1938.37				0.56			
Obs.	424	424	424	424	780	780	780	780

Notes: OLS regression based on data from Abebe et al. (2020). The dependent variable in the first four columns is the natural logarithm of wages in the first endline survey. The dependent variable in the last four columns is a dummy equal to one if the respondent is employed at the time of the first endline survey. 'Raven' is the number of correct answers in a 60 item Raven test, which was administered shortly after the baseline interview. 'Conscientiousness' and 'neuroticism' are the conscientiousness and neuroticism scores obtained by administering a 10-items BFI inventory at baseline. All ability variables are standardised so that they have a mean of zero and a standard deviation of one. We report the mean of the dependent variables in the second-to-last row. For the wage variable, we report the mean wage in Ethiopian Birr units (as opposed to the mean of the natural logarithm).

Table A.3: Ability and labour market outcomes with controls

Dep. var.	Ln(wage)				Employed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven	.084			.084	.048			.047
	(0.034)			(0.034)	(0.018)			(0.018)
Conscientiousness	-.032	.015		.014		.007		.004
	(0.025)	(0.030)		(0.032)		(0.018)		(0.019)
Neuroticism	-.047		-.001	0			-.008	-.007
	(0.025)		(0.034)	(0.035)			(0.018)	(0.019)
Mean dep. var.	1938.37				0.56			
Obs.	424	424	424	424	780	780	780	780

Notes: OLS regression based on data from Abebe et al. (2020). The dependent variable in the first four columns is the natural logarithm of wages in the first endline survey. The dependent variable in the last four columns is a dummy equal to one if the respondent is employed at the time of the first endline survey. 'Raven' is the number of correct answers in a 60 item Raven test, which was administered shortly after the baseline interview. 'Conscientiousness' and 'neuroticism' are the conscientiousness and neuroticism scores obtained by administering a 10-items BFI inventory at baseline. All ability variables are standardised so that they have a mean of zero and a standard deviation of one. We report the mean of the dependent variables in the second-to-last row. For the wage variable, we report the mean wage in Ethiopian Birr units (as opposed to the mean of the natural logarithm). All regressions include controls for age and age squared and a dummy for having worked in a permanent job in the past.

Table A.4: The probability of finding a job of high and low-cost individuals

	Search-to-work transition	
	(1)	(2)
Raven (z-score)	.022	.016
	(0.019)	(0.018)
Low saving * Raven	-.047	
	(0.025)	
High distance * Raven		-.038
		(0.026)
Low saving	-.032	
	(0.025)	
High distance		.016
		(0.025)
Constant	.235	.207
	(0.020)	(0.016)
Obs.	2218	2218

Notes: OLS regression based on data from Abebe et al. (2020). The data is collapsed at the monthly level. We restrict the sample to unemployed people who are searching for work. The dependent variable is a dummy capturing whether the respondent is employed in at least one of the fortnights of the following month. The model enables to estimate the correlation between the probability of finding a job in the following month and (i) the z-score of the Raven test, (ii) a proxy for application costs, and (iii) the interaction between the two. Standard errors clustered at the individual level reported in parenthesis.

Table A.5: Psychometrics Validity Checks: Cronbach α

	Raw	Ipsatised	Laajaj and Macours (2017)
Conscientiousness	0.59	0.70	.51
Neuroticism	0.61	0.62	.31
Grit	0.59	0.72	

Notes: In the first column, we report the value of Cronbach α for our three main measures of non-cognitive ability. In the second column, we show the Cronbach α for the ipsatised values of these variables. Variables are ipsatised by subtracting the individual acquiescence score — the mean of positive items and inverted items across scales — which is a measure of the tendency to agree with any statement (Laajaj and Macours, 2017). In third column, we report for reference the values of Cronbach α from a recent study in Kenya by Laajaj and Macours (2017). Sample used: all applicants.

Table A.6: Indices of applicant quality

Index	Variable	Measure
Cognitive ability	Raven	No. of correct answers
	Stroop	Time in seconds
	Stroop	No. mistakes
Non-cognitive ability	Conscientiousness	BFI44 score
	Neuroticism	BFI44 score
	Grit	Score on grit scale
Experience	Routine tasks	No. months
	Managerial tasks	No. months
	Problem solving tasks	No. months

Table A.7: Correlation between indices

	Cognitive ability	Non-cognitive ability	Experience
Cognitive ability	1		
Non-cognitive ability	0.205	1	
Experience	-0.002	0.064	1

Notes: Correlation coefficients. Sample used: all applicants.

Table A.8: Summary statistics and balance for the sample of non-attriters

	Mean			StDev	N	Balance tests (p)		
	Control	Incentive	High wage	Control		Incentive = Wage = Control	Incentive = Control	Wage = Control
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.21	0.21	0.21	0.41	4386	0.90	0.71	0.68
Age	25.96	25.76	26.10	4.23	4383	0.11	0.20	0.41
Born in Addis Ababa	0.24	0.24	0.23	0.43	4386	0.59	0.81	0.45
First language is Amharic	0.68	0.70	0.67	0.47	4386	0.20	0.16	0.81
Heard about job in newspaper	0.55	0.58	0.55	0.50	4386	0.21	0.12	0.97
Engineering or hard science	0.51	0.49	0.49	0.50	4386	0.44	0.34	0.22
Economics	0.15	0.17	0.16	0.36	4386	0.62	0.34	0.54
Other social science	0.15	0.15	0.14	0.36	4386	0.45	0.68	0.41
Wage work experience (dummy)	0.52	0.51	0.52	0.50	4386	0.91	0.83	0.83
Wage work experience (months)	27.15	26.93	28.27	42.92	4386	0.71	0.89	0.51
Self-employed experience	0.33	0.34	0.35	0.47	4386	0.67	0.61	0.37
Currently unemployed	0.67	0.66	0.64	0.47	4386	0.15	0.33	0.05
Currently wage employed	0.24	0.25	0.27	0.43	4386	0.18	0.45	0.07
GPA	2.98	2.99	2.96	0.46	4168	0.20	0.48	0.28
Overall balance						0.24	0.53	0.38

Notes: Balance table. In columns 1-5, we report descriptive statistics and sample sizes for the sample of individuals that were interviewed in the second phone survey. In columns 6-8, we report the p-values of a battery of balance tests. In column 6, we test the null hypothesis that three experimental groups are balanced. In columns 7 and 8, we present pair-wise tests of a single treatment group against the control. In the last row, we report a joint test of orthogonality (following the recent literature, e.g. McKenzie (2017b)). To perform the joint test of orthogonality we regress the treatment variable on all covariates and we then test the joint null hypothesis that all covariates have a zero coefficient. We do this by estimating a categorical logit model in column 6 (the treatment variable can take three values) and OLS models in columns 7 and 8 (here we drop one experimental group, so our treatment variable only takes two values). Sample used: second phone call sample.

Table A.9: Additional balance tests

	Week							
	1	2	3	4	5	6	7	8
Incentive	-.001	0	0	.001	0	.002	-.001	-.002
	(0.010)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)
High Wage	.001	-.001	-.002	.002	-.001	.001	0	0
	(0.010)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)

Notes: OLS regression testing balance of the day of the week when the application test took place. Robust standard errors reported in parenthesis. Sample used: baseline sample.

Table A.10: Is the sample representative?

	Age	Female	Work experience	Unemployment duration
Experiment	25.13	.23	.34	4.99
	(3.85)	(0.42)	(0.47)	(7.02)
Labour Force Survey	26.72	.36	.29	9.16
	(8.38)	(0.49)	(0.46)	(7.52)

Notes: This table presents means and standard deviations of key variables in the experimental sample and in a representative sample of jobseekers from Addis Ababa who (i) use job boards or newspapers for job search and (ii) have the educational qualifications required to apply for the experiment's job (they hold a vocational diploma or a university degree). Unemployment duration is measured in months. Sample used: baseline sample and 2013 Labour Force Survey sample.

Table A.11: Components of index

	Raven	Stroop time	Stroop mistakes
	(1)	(2)	(3)
Incentive	1.155	-2.601	-0.050
	(0.618)	(1.046)	(0.188)
	[0.092]	[0.039]	[0.791]
High wage	0.591	-0.982	-0.302
	(0.618)	(1.046)	(0.188)
	[0.337]	[0.337]	[0.297]
Control mean	38.593	117.304	3.854
Incentive = Wage (p)	0.307	0.078	0.098
Obs.	2397	2386	2388

Notes: OLS regression. The dependent variable is indicated in the column heading. 'Raven' is the number of correctly answered questions on the Raven test. 'Stroop time' is the number of seconds required to complete the Stroop test. 'Stroop mistakes' is the number of mistakes made in the Stroop task. The negative coefficients on 'Stroop time' and 'Stroop mistakes' indicate better performance. Sharpened q-values (Benjamini et al., 2006) reported in brackets. q-values control the false discovery rate for the multiple comparisons reported in the same row of the table. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.12: Impacts on applicant GPA

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.049	.07	.05	.06	.04	.08
	(0.025)	(0.043)	(0.035)	(0.033)	(0.033)	(0.036)
High Wage	.012	-.01	-.01	.03	0	.04
	(0.024)	(0.040)	(0.033)	(0.033)	(0.030)	(0.039)
Control value	2.94	3.56	3.27	2.93	2.60	2.32
Incentive = Wage (p)	0.088	0.022	0.045	0.360	0.181	0.187
Obs.	2285	2285	2285	2285	2285	2285

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The dependent variable is the applicant's GPA. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.13: Proportion of applicants who score above a threshold

	Threshold (percentile in control group distribution)				
	90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)
Incentive	.053	.229	.229	.17	.412
	(0.024)	(0.110)	(0.117)	(0.133)	(0.173)
High Wage	.052	.202	.227	.075	.28
	(0.023)	(0.108)	(0.112)	(0.131)	(0.165)
Incentive = Wage (p)	0.966	0.795	0.983	0.448	0.371
Obs.	2386	2386	2386	2386	2386

Notes: OLS regression. The dependent variable is a dummy for whether the applicant's cognitive ability is above the threshold indicated in the column heading. Thresholds are defined with respect to specific percentiles of the control distribution of cognitive ability. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.14: Average ability of top candidates

	Top 20		Top 10		Top 5	
	Cognitive	GPA	Cognitive	GPA	Cognitive	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.359	.102	.243	.182	.15	.233
	(0.078)	(0.050)	(0.091)	(0.069)	(0.114)	(0.095)
High Wage	.457	.066	.354	.1	.282	.144
	(0.078)	(0.052)	(0.092)	(0.074)	(0.116)	(0.100)
Control group st. dev.	0.71	0.45	0.59	0.45	0.53	0.42
Incentive = Wage (p)	0.173	0.480	0.206	0.243	0.260	0.324
Obs.	480	466	240	233	120	116

Notes: Estimates from OLS regressions. The sample comprises the top 20, 10 and 5 applicants for job offered (top applicants are defined using the cognitive ability score, following the procedures used in the field experiment). The employer offers one job per fortnight for each treatment group. In a few cases, the employer combines two fortnights for the same treatment group and offers only one job for these two fortnights. For the present analysis, however, we consider the top applicants from each fortnight separately. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all top applicants.

Table A.15: Non-cognitive ability

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.095	-.155	-.161	-.004	-.141	-.137
	(0.125)	(0.134)	(0.132)	(0.157)	(0.199)	(0.229)
	[0.566]	[0.369]	[0.336]	[1.000]	[0.717]	[1.000]
High Wage	.17	0	-.039	.257	.162	.26
	(0.118)	(0.127)	(0.129)	(0.147)	(0.203)	(0.222)
	[0.225]	[1.000]	[0.764]	[0.245]	[0.637]	[0.725]
Control value	0.0000	2.688	1.721	0.241	-1.227	-2.885
Incentive = Wage (p)	0.015	0.235	0.288	0.038	0.095	0.091
Obs.	2373	2373	2373	2373	2373	2373

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The dependent variable is the index of non-cognitive ability. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) are reported in brackets. q-values control the false discovery rate for the multiple tests of the same hypothesis for different indices of ability. A Wilcoxon rank-sum test fails to reject the equality of the distribution of non-cognitive ability in the control and incentive groups (p=.411) and fails to reject the equality of the distribution of non-cognitive ability in the control and wage groups (p=.255). Sample used: all applicants.

Table A.16: Experience

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.091	.163	-.044	0	0	0
	(0.158)	(0.850)	(0.117)	(0.008)	(0.004)	(0.002)
	[0.566]	[0.848]	[0.704]	[1.000]	[1.000]	[1.000]
High Wage	-.063	.302	-.088	0	0	0
	(0.157)	(0.733)	(0.147)	(0.008)	(0.004)	(0.002)
	[0.687]	[1.000]	[0.764]	[1.000]	[1.000]	[1.000]
Control value	0.0000	2.217	0.064	-1.225	-1.225	-1.225
Incentive = Wage (p)	0.808	0.811	0.718	1.000	1.000	1.000
Obs.	2311	2311	2311	2311	2311	2311

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The dependent variable is the experience index. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) are reported in brackets. q-values control the false discovery rate for the multiple tests of the same hypothesis for different indices of ability. A Wilcoxon rank-sum test fails to reject the equality of the distribution of experience in the control and incentive groups (p=.354) and fails to reject the equality of the distribution of experience in the control and wage groups (p=.718). We report results obtained using an alternative definition of the index in Table . used: all applicants.

Table A.17: Cognitive ability, weighted index

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.079	.068	.06	.062	.114	.022
	(0.037)	(0.035)	(0.038)	(0.045)	(0.060)	(0.074)
High Wage	.066	.065	.067	.028	.085	.037
	(0.037)	(0.035)	(0.036)	(0.043)	(0.057)	(0.070)
Control value	-0.0000	0.769	0.499	0.113	-0.403	-0.864
Incentive = Wage (p)	0.699	0.917	0.819	0.396	0.575	0.847
Obs.	2386	2386	2386	2386	2386	2386

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. Cognitive ability index obtained by weighting observations by the inverse of the covariance matrix. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis. A Wilcoxon rank-sum test rejects the equality of the distribution of cognitive ability in the control and incentive groups (p=.057) and marginally rejects the equality of the distribution of cognitive ability in the control and wage groups (p=.096). Sample used: all applicants.

Table A.18: Non cognitive ability, weighted index

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.036	-.059	-.069	.015	-.068	-.051
	(0.042)	(0.050)	(0.043)	(0.052)	(0.068)	(0.091)
High Wage	.057	.003	-.035	.102	.068	.109
	(0.039)	(0.047)	(0.043)	(0.050)	(0.065)	(0.082)
Control value	0.0000	0.929	0.585	0.059	-0.423	-0.988
Incentive = Wage (p)	0.010	0.177	0.387	0.049	0.018	0.053
Obs.	2373	2373	2373	2373	2373	2373

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. Non-cognitive ability index obtained by weighting observations by the inverse of the covariance matrix. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis. A Wilcoxon rank-sum test fails to reject the equality of the distribution of non-cognitive ability in the control and incentive groups (p=.371) and fails to reject the equality of the distribution of non-cognitive ability in the control and wage groups (p=.239). Sample used: all applicants.

Table A.19: Experience, weighted index

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.029	.007	-.014	0	0	0
	(0.053)	(0.142)	(0.058)	(0.002)	(0.001)	(0.000)
High Wage	-.02	.063	-.022	0	0	0
	(0.052)	(0.127)	(0.057)	(0.002)	(0.001)	(0.000)
Control value	-0.0000	0.753	0.017	-0.409	-0.409	-0.409
Incentive = Wage (p)	0.820	0.685	0.830	1.000	1.000	1.000
Obs.	2311	2311	2311	2311	2311	2311

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. Experience ability index obtained by weighting observations by the inverse of the covariance matrix. The second-to-last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors are reported in parenthesis (for the 50th, 25th and 10th we estimate coefficients of zero and are thus unable to calculate robust standard errors; for these percentiles we report raw standard errors instead). A Wilcoxon rank-sum test fails to reject the equality of the distribution of experience in the control and incentive groups (p=.354) and fails to reject the equality of the distribution of experience in the control and wage groups (p=.719). Sample used: all applicants.

Table A.20: Non-cognitive ability, ipsatised

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.113	-.2	-.213	.024	-.075	-.174
	(0.118)	(0.141)	(0.126)	(0.154)	(0.178)	(0.226)
High Wage	.09	-.082	-.092	.18	.166	.228
	(0.112)	(0.138)	(0.120)	(0.142)	(0.172)	(0.224)
Control value	-0.00	2.68	1.60	0.15	-1.28	-2.83
Incentive = Wage (p)	0.050	0.325	0.278	0.221	0.166	0.051
Obs.	2373	2373	2373	2373	2373	2373

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The dependent variable is an index of ipsatised non-cognitive ability. The index is based on ipsatised values of conscientiousness, neuroticism and grit. Variables are ipsatised by subtracting the individual acquiescence score — the mean of positive items and inverted items across scales — which is a measure of the tendency to agree with any statement (Laajaj and Macours, 2017). The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. A Wilcoxon rank-sum test fails to reject the equality of the distribution of ipsatised non-cognitive ability in the control and incentive groups (p=.354) and fails to reject the equality of the distribution of ipsatised non-cognitive ability in the control and wage groups (p=.466). Sample used: all applicants.

Table A.21: Job search outcomes in 30 days after first phone call

	Applications	Money (USD)	Time	Interviews	Offers	Has job
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.009	.055	27.159	-.011	-.013	-.002
	(0.072)	(0.220)	(25.206)	(0.024)	(0.013)	(0.007)
High Wage	-.077	.012	-11.817	-.039	-.028	-.019
	(0.066)	(0.216)	(23.438)	(0.023)	(0.012)	(0.007)
Control group mean	1.573	2.742	392.509	.309	.103	.048
Incentive = Wage (p)	0.313	0.852	0.107	0.231	0.233	0.012
Obs.	4328	4328	4328	4328	4328	4328

Notes: OLS regression. The dependent variable is indicated in the column heading. All dependent variables are measured in the second phone call (see Figure A.2 for a timeline). Further, all dependent variables refer to jobs other than the experiment's job. These variables are collected through a short application roster where the respondent is asked a number of questions about each application they have made in the 30 days between the two phone calls. This includes information about the application process and its outcome. The variables 'applications', 'interviews' and 'offers' capture the total number of applications and interviews made, and offers received. 'Has job' is a dummy variable capturing whether the respondent is currently employed in one of the jobs they have applied for in the period between the two phone calls. 'Cost' and 'time' are, respectively, the total amount of money and time that the respondent reports to have spent on all job applications they have made in the 30 days period. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.22: Additional job search outcomes in 30 days after first phone call (quality of applications)

	Number of applications				
	Total (incl. exp. job)	Occup. matched	Skills matched	Long run	Permanent
	(1)	(2)	(3)	(4)	(5)
Incentive	.114	.015	-.029	-.034	.011
	(0.074)	(0.055)	(0.031)	(0.039)	(0.055)
High Wage	.113	-.036	-.066	.004	-.033
	(0.069)	(0.054)	(0.031)	(0.040)	(0.054)
Control group mean	1.983	1.299	.357	.502	1.309
Incentive = Wage (p)	0.990	0.346	0.226	0.322	0.417
Obs.	4328	4328	4328	4328	4328

Notes: OLS regression. The dependent variable is indicated in the column heading. All dependent variables are measured in the second phone call (see Figure A.2 for a timeline). 'Applications' is the total number of applications made. This variable includes the application to the experiment's job and thus differs from the variable 'Applications' reported in Table A.21. On the other hand, the variables reported in columns (2)-(5) do not include the application to the experiment's job. For each application made to these other jobs, respondents are asked a number of questions about the position: the occupation of the job, whether they feel they have the right skills for the job, a rating from 0 to 10 indicating whether they see themselves doing that particular job in the long run (we create a dummy that splits this variable at the median), and whether the job has an open-ended contract. We use these responses to construct three variables: (2) 'Occup. matched' is the number of applications to positions that match the occupation the jobseeker would like to find, (3) 'Skills matched' is the number of applications to positions that match the skills of the jobseeker (i.e. the jobseeker does not feel overqualified for the position) (4) 'Long run' is the number of applications to positions that the jobseeker sees herself doing in the long run and (5) 'Permanent' is the number of applications to positions that offer an open-ended contract. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.23: Additional job search outcomes in 30 days after first phone call (quality of interviews and jobs)

	Interviews for jobs that are...			Job is...		
	Matched	Long run	Permanent	Matched	Long run	Permanent
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.015	-.023	0	-.004	-.006	-.004
	(0.022)	(0.015)	(0.022)	(0.007)	(0.005)	(0.007)
High Wage	-.04	-.023	-.028	-.018	-.013	-.02
	(0.022)	(0.015)	(0.022)	(0.007)	(0.005)	(0.007)
Control group mean	.273	.129	.261	.045	.024	.043
Incentive = Wage (p)	0.245	0.980	0.209	0.031	0.154	0.012
Obs.	4328	4328	4328	4328	4328	4328

Notes: OLS regression. The dependent variable is indicated in the column heading. All dependent variables are measured in the second phone call (see Figure A.2 for a timeline). The first set of dependent variables are defined as the number of interviews for jobs that: (1) match the occupation the jobseeker would like to find, (2) the jobseeker sees herself doing in the long run and (3) offer an open-ended contract. The second set of dependent variables are defined as a dummy variable for working in a job that (4) matches the occupation the jobseeker would like to find, (5) the respondent sees herself doing in the long run and (6) offers an open-ended contract. All variables exclude the application to the experiment's job. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.24: Search method

	Board	Newspaper	Direct	Network
	(1)	(2)	(3)	(4)
Incentive	0.037	-0.032	-0.002	-0.054
	(0.306)	(0.162)	(0.064)	(0.199)
High Wage	-0.624	-0.152	-0.049	-0.064
	(0.294)	(0.161)	(0.061)	(0.235)
Control Mean	6.326	3.895	0.567	2.272
St. dev.	8.077	4.365	1.725	6.016
Incentive = Wage (p)	0.028	0.460	0.436	0.963
Obs.	4357	4366	4370	4349

Notes: OLS regressions. The dependent variable is reported in the column headings. 'Board' is the number of days in the last 30 days when the respondent visited the job vacancy board. 'Newspaper' is the number of times in the last 30 days when the respondent consulted the job insert in the newspaper. 'Direct' is the number of days in the last 30 days when the respondent visited employers to inquire about vacancies. 'Network' is the number of social contacts that the person has talked to about job opportunities in the last 30 days. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.25: Heterogeneous impacts on applications

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	.073	.092	.097	.108	.101	.098
	(0.039)	(0.025)	(0.021)	(0.033)	(0.027)	(0.028)
High wage	.19	.217	.214	.242	.196	.206
	(0.038)	(0.024)	(0.021)	(0.032)	(0.027)	(0.026)
Control mean	.429	.492	.473	.464	.464	.504
Incentive = Wage (p)	0.002	0.000	0.000	0.000	0.000	0.000
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.125	.136	.158	.081	.121	.132
	(0.020)	(0.025)	(0.030)	(0.030)	(0.023)	(0.025)
High wage	.185	.156	.147	.205	.179	.172
	(0.020)	(0.025)	(0.030)	(0.029)	(0.023)	(0.025)
Control mean	.407	.329	.27	.482	.372	.317
Incentive = Wage (p)	0.003	0.436	0.704	0.000	0.015	0.112
No het. incentive (p)	0.235	0.209	0.098	0.536	0.567	0.329
No het. wage (p)	0.891	0.080	0.065	0.383	0.631	0.322
Obs.	4689	4686	4689	3020	4686	4686

Notes: OLS regressions. The dependent variable is a dummy capturing whether the jobseeker applied for the experiment's job. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: baseline sample.

Table A.26: Heterogeneous impacts on other job search

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	-.014	.065	.021	.068	-.01	.025
	(0.142)	(0.120)	(0.105)	(0.166)	(0.120)	(0.140)
High wage	-.128	-.065	-.058	.021	-.174	-.125
	(0.137)	(0.097)	(0.092)	(0.129)	(0.117)	(0.115)
Control mean	1.407	1.811	1.779	1.472	1.78	1.801
Incentive = Wage (p)	0.383	0.259	0.405	0.762	0.110	0.193
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.005	-.073	-.027	.015	.002	-.021
	(0.094)	(0.102)	(0.100)	(0.141)	(0.107)	(0.094)
High wage	-.083	-.129	-.107	-.06	-.025	-.048
	(0.083)	(0.105)	(0.101)	(0.136)	(0.089)	(0.096)
Control mean	1.617	1.322	1.093	2.046	1.41	1.332
Incentive = Wage (p)	0.300	0.499	0.398	0.547	0.791	0.754
No het. incentive (p)	0.908	0.381	0.738	0.805	0.940	0.767
No het. wage (p)	0.781	0.653	0.717	0.666	0.314	0.590
Obs.	4328	4325	4328	2804	4325	4325

Notes: OLS regressions. The dependent variable is the number of applications to jobs other than the experiment's job. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: second phone call sample.

Table A.27: Heterogeneous impacts on cognitive ability

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	1.153	.428	.377	.548	.411	.425
	(0.270)	(0.155)	(0.129)	(0.217)	(0.169)	(0.158)
High wage	.447	.354	.272	.377	.244	.332
	(0.272)	(0.152)	(0.127)	(0.216)	(0.171)	(0.155)
Control mean	-.317	-.08	.011	-.34	.03	-.025
Incentive = Wage (p)	0.000	0.562	0.343	0.308	0.241	0.461
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.008	-.006	-.116	.217	.096	-.004
	(0.121)	(0.154)	(0.214)	(0.165)	(0.148)	(0.166)
High wage	.123	-.044	-.064	.215	.152	-.006
	(0.118)	(0.152)	(0.214)	(0.156)	(0.142)	(0.166)
Control mean	.089	.124	-.044	.281	-.028	.041
Incentive = Wage (p)	0.295	0.798	0.794	0.990	0.672	0.993
No het. incentive (p)	0.000	0.047	0.049	0.225	0.161	0.050
No het. wage (p)	0.274	0.064	0.175	0.544	0.679	0.118
Obs.	2386	2385	2386	1738	2384	2384

Notes: OLS regressions. The dependent variable is the cognitive ability index. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: all applicants.

Table A.28: Heterogeneous impacts on Raven score

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	4.791	2.291	1.894	1.666	2.758	2.162
	(1.306)	(0.819)	(0.699)	(1.077)	(0.899)	(0.822)
High wage	0.966	1.311	0.932	0.141	1.387	0.850
	(1.349)	(0.798)	(0.680)	(1.035)	(0.906)	(0.815)
Control mean	37.179	38.765	38.912	38.281	38.676	39.407
Incentive = Wage (p)	0.001	0.164	0.118	0.106	0.085	0.061
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.188	-.276	-.621	1.778	-.315	.002
	(0.696)	(0.919)	(1.270)	(0.959)	(0.841)	(0.985)
High wage	.49	-.422	-.335	1.377	-.097	.416
	(0.678)	(0.925)	(1.312)	(0.926)	(0.813)	(0.963)
Control mean	38.99	38.327	37.307	39.582	38.515	37.267
Incentive = Wage (p)	0.632	0.868	0.809	0.638	0.774	0.633
No het. incentive (p)	0.002	0.037	0.083	0.938	0.013	0.081
No het. wage (p)	0.752	0.156	0.391	0.373	0.223	0.724
Obs.	2397	2396	2397	1743	2395	2395

Notes: OLS regressions. The dependent variable is the Raven test score. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: all applicants.

Table A.29: Heterogeneous impacts on GPA

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	.129	.078	.067	.083	.064	.077
	(0.053)	(0.029)	(0.027)	(0.041)	(0.034)	(0.033)
High wage	0.081	0.067	0.036	0.061	0.021	0.048
	(0.050)	(0.029)	(0.027)	(0.040)	(0.034)	(0.033)
Control mean	2.797	2.987	2.973	2.897	3.027	3
Incentive = Wage (p)	0.359	0.672	0.223	0.545	0.162	0.277
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.024	.009	-.01	.047	.035	.011
	(0.027)	(0.043)	(0.055)	(0.038)	(0.035)	(0.044)
High wage	-.008	-.073	-.075	.018	.004	-.039
	(0.026)	(0.041)	(0.054)	(0.036)	(0.033)	(0.040)
Control mean	3.013	2.949	2.959	3.019	2.925	2.935
Incentive = Wage (p)	0.178	0.028	0.161	0.405	0.314	0.168
No het. incentive (p)	0.078	0.182	0.212	0.513	0.555	0.200
No het. wage (p)	0.118	0.005	0.063	0.425	0.728	0.082
Obs.	2285	2284	2285	1670	2283	2283

Notes: OLS regressions. The dependent variable is the applicant's GPA. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: all applicants.

Table A.30: Heterogeneous impacts on top applicants (75th percentile)

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Low experience	Unemployed	Long spell	Young	High value
Incentive	.24	.081	.073	.077	.094	.074
	(0.052)	(0.031)	(0.027)	(0.040)	(0.035)	(0.038)
High wage	.097	.061	.057	.06	.063	.048
	(0.047)	(0.030)	(0.026)	(0.038)	(0.034)	(0.037)
Control top applicants	.209	.259	.259	.216	.27	.274
Incentive = Wage (p)	0.004	0.494	0.535	0.653	0.351	0.372
Impacts for	Male	High experience	Employed	Short spell	Old	Low value
Incentive	.003	.018	.005	.065	.014	.029
	(0.026)	(0.035)	(0.046)	(0.039)	(0.031)	(0.038)
High wage	.039	.04	.041	.057	.044	.064
	(0.026)	(0.035)	(0.047)	(0.037)	(0.031)	(0.038)
Control top applicants	.26	.233	.206	.295	.229	.207
Incentive = Wage (p)	0.142	0.503	0.377	0.828	0.311	0.276
No het. incentive (p)	0.000	0.187	0.204	0.834	0.090	0.333
No het. wage (p)	0.279	0.661	0.770	0.956	0.670	0.723
Obs.	2386	2385	2386	1738	2384	2384

Notes: OLS regressions. The dependent variable is a dummy capturing whether the jobseeker scored above the 75th percentile of the control distribution of the cognitive ability index. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: all applicants.

Table A.31: Heterogeneous impacts on quality at the top of the distribution

Heterogeneity by	Gender		Experience		Unemployed		Length unemployment		Age		Value job	
	p90	p75	p90	p75	p90	p75	p90	p75	p90	p75	p90	p75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Impacts for	Female		Low experience		Unemployed		Long spell		Young		High value	
Incentive	.582	.848	.244	.176	.227	.185	.259	.29	.131	.131	.186	.112
	(0.241)	(0.230)	(0.136)	(0.155)	(0.122)	(0.134)	(0.188)	(0.204)	(0.161)	(0.168)	(0.148)	(0.157)
High wage	.247	.257	.165	.136	.136	.139	.146	.257	.164	.103	.117	.029
	(0.246)	(0.225)	(0.129)	(0.153)	(0.120)	(0.129)	(0.205)	(0.178)	(0.168)	(0.175)	(0.146)	(0.150)
Control quality	-.317	2.267	-.08	2.404	.011	2.421	-.34	2.223	.03	2.535	-.025	2.442
Incentive = Wage (p)	0.074	0.002	0.534	0.756	0.416	0.682	0.519	0.870	0.821	0.835	0.583	0.484
Impacts for	Male		High experience		Employed		Short spell		Old		Low value	
Incentive	.081	-.069	.252	.124	.283	.033	.104	.085	.266	.086	.345	.178
	(0.126)	(0.131)	(0.189)	(0.168)	(0.217)	(0.258)	(0.208)	(0.179)	(0.160)	(0.150)	(0.157)	(0.182)
High wage	.175	.126	.098	.262	.167	.189	.045	.118	.143	.244	.255	.336
	(0.121)	(0.130)	(0.185)	(0.161)	(0.222)	(0.252)	(0.192)	(0.173)	(0.152)	(0.139)	(0.180)	(0.171)
Control quality	.089	2.339	.124	2.211	-.044	2.059	.281	2.654	-.028	2.211	.041	2.114
Incentive = Wage (p)	0.420	0.075	0.320	0.358	0.549	0.513	0.738	0.822	0.408	0.260	0.576	0.338
No het. incentive (p)	0.066	0.001	0.973	0.821	0.821	0.602	0.580	0.449	0.553	0.842	0.501	0.777
No het. wage (p)	0.793	0.614	0.768	0.572	0.903	0.860	0.719	0.576	0.927	0.528	0.533	0.186
Obs.	2386	2386	2385	2385	2386	2386	1738	1738	2384	2384	2384	2384

Notes: Quantile regressions. The dependent variable is the cognitive ability score at a given percentile of the control distribution (90th or 75th). Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of a test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. The standard errors reported in column (6) are bootstrapped to reflect the uncertainty in the estimation of the present value of the job. Sample used: all applicants.

Table A.32: Decomposition of impact on cognitive ability

Heterogeneity by	Gender	Experience	Unemployed	Length unemployment	Age	Value job
	(1)	(2)	(3)	(4)	(5)	(6)
Compositional effect	0.03	0.04	-0.01	-0.06	-0.00	0.01
Within-group effect for	Female	Low experience	Unemployed	Long spell	Young	High value
	0.95	0.97	1.13	0.73	0.79	0.99
Within-group effect for	Male	High experience	Employed	Short spell	Old	Low value
	0.03	-0.01	-0.12	0.33	0.21	-0.01

Notes: Decomposition of the treatment effect reported in Table A.27. We implement this decomposition as follows. There are six dimensions of heterogeneity, namely gender, experience, unemployment status, unemployment length, age, and job value. Each of these dimensions is split in two categories (male/female, etc.). We denote these with a vector of dummy variables $v_i \in \{0,1\}$ for all dimensions $i = \{1,2,...,6\}$. Further, we use $j \in \{0,1\}$ to indicate the experimental group ($j = 0$ refers to the control group and $j = 1$ to the incentive group). Finally, we use p_i^j to indicate the share of applicants in group j for whom $v_i = 1$. For each dimension of heterogeneity i, we decompose the total effect on expected ability T into three components: (i) a compositional effect: $(p_i^1 - p_i^0) * (E[T|v_i = 1, j = 0] - E[T|v_i = 0, j = 0])$; (ii) a within-group effect for the first group of applicants: $p_i^1 * (E[T|v_i = 1, j = 1] - E[T|v_i = 1, j = 0])$; (iii) a within-group effect for the second group of applicants: $(1 - p_i^1) * (E[T|v_i = 0, j = 1] - E[T|v_i = 0, j = 0])$.

.

Table A.33: Additional heterogeneity: first language

	Apply	Cognitive	Raven	GPA	No. other applications
	(1)	(2)	(3)	(4)	(5)
Impacts for			First language Amharic		
Incentive	0.123	0.127	0.708	0.042	0.075
	(0.021)	(0.129)	(0.734)	(0.020)	(0.101)
High Wage	0.210	0.138	0.251	-0.007	-0.091
	(0.021)	(0.126)	(0.733)	(0.020)	(0.086)
Control Mean	0.390	0.384	40.140	2.946	1.618
Incentive = Wage (p)	0.000	0.920	0.474	0.017	0.067
Impacts for			First language not Amharic		
Incentive	0.098	0.360	1.527	-0.016	-0.188
	(0.032)	(0.202)	(1.101)	(0.031)	(0.125)
High Wage	0.136	0.196	0.811	-0.002	-0.099
	(0.031)	(0.200)	(1.045)	(0.030)	(0.130)
Control Mean	0.457	-0.679	35.861	3.017	1.482
Incentive = Wage (p)	0.231	0.372	0.492	0.640	0.437
No het. incentive (p)	0.524	0.330	0.536	0.117	0.101
No het. wage (p)	0.049	0.805	0.661	0.895	0.964
Obs.	4689	2386	2397	4451	4328

Notes: OLS regressions. The dependent variable is reported in the column headings. Robust standard errors reported in parenthesis. Sample used: baseline sample (column 1), all applicants (columns 2-4), second phone call sample (column 5).

Table A.34: Time preferences and cost of effort

	β	δ	γ
	(1)	(2)	(3)
Control	0.817	0.740	5.842
	(0.083)	(0.090)	(0.784)
Incentive	0.794	0.891	5.032
	(0.057)	(0.064)	(0.500)
High Wage	0.811	0.971	5.233
	(0.053)	(0.070)	(0.511)
Incentive - Control (p)	0.825	0.420	0.384
Incentive - Wage (p)	0.833	0.400	0.778
Wage - Control (p)	0.953	0.044	0.515

Notes: Structural estimates of present bias (β), impatience (δ) and cost of effort (γ). The estimation technique is described in detail in Appendix A.3. Standard errors obtained with the delta method reported in parenthesis. The last three rows report the p-values of tests of the equality of the coefficients. Sample used: all applicants.

Table A.35: Preferences and sophistication

	Present bias	Risk Preferences	Social Preferences	Sophistication
	(1)	(2)	(3)	
Incentive	.022	.054	-.033	.007
	(0.027)	(0.066)	(0.097)	(0.021)
High Wage	.005	.03	.042	.001
	(0.025)	(0.064)	(0.094)	(0.020)
Variable	β_i	Index	Index	Dummy
Control group mean	.301	1.905	.04	.19
Control group st.dev.	.459	1.17	1.762	.393
Incentive = Wage (p)	0.466	0.684	0.397	0.751
Obs.	2053	2110	2193	2331

Notes: Estimates from OLS regressions. Robust standard errors reported in parenthesis. Present bias is a dummy for individuals with $\beta < 0.99$. Sophistication is a dummy for individuals with $k \geq 2$. The tasks used to elicit these variables are described in detail in Appendix A.3. Sample size changes because of missing responses and because we are not able to estimate a β coefficient for all choice patterns. Sample used: all applicants.

Table A.36: Test effort

	Mistake		Unfinished	
	(1)	(2)	(3)	(4)
Incentive	.096	.033	-.019	-.005
	(0.082)	(0.025)	(0.038)	(0.007)
High Wage	.067	.021	-.028	-.002
	(0.078)	(0.024)	(0.034)	(0.007)
Measure	continuous	dummy	continuous	dummy
Control group mean	.711	.292	.081	.018
Incentive = Wage (p)	0.712	0.569	0.782	0.701
Obs.	2316	2316	2332	2332

Notes: OLS regression. In column (1), the dependent variable 'Mistake' is the number of strings transcribed incorrectly. In column (2), the dependent variable is a dummy capturing whether any string was transcribed incorrectly. In column (3), the dependent variable 'Unfinished' is the number of strings that the applicant has failed to transcribe. In column (4), the dependent variable is a dummy capturing whether the applicant has failed to transcribe any string. Robust standard errors reported in parenthesis. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Sample used: all applicants.

Table A.37: Labour market behaviour of applicants

	No. other applications	Rejected offer (dummy)
	(1)	(2)
Incentive	.075	-.029
	(0.082)	(0.038)
High Wage	.053	.138
	(0.116)	(0.103)
Control group mean	1.677	.439
Incentive = Wage (p)	0.806	0.073
Obs.	2215	154

Notes: OLS regression for the sample of individuals who have applied to the experiment's job. The dependent variable is indicated in the column headings. 'No. applicants' is the number of applications to other jobs that the individual made between the two phone calls. 'Rejected (dummy)' is a dummy for whether the individual rejected a job offer (which is set to missing for individuals who did not receive a job offer in this time period). The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.38: Salience of the position

	Correct answer		Absolute mistake	
	(1)	(2)	(3)	(4)
Incentive	.025	.024	-31.22	-28.378
	(0.015)	(0.016)	(18.754)	(19.221)
High Wage	.039	.038	-45.408	-40.252
	(0.015)	(0.016)	(18.095)	(19.421)
Control for application	no	yes	no	yes
Control group mean	0.686	0.686	167.303	167.303
Incentive = Wage (p)	0.349	0.370	0.362	0.453
Obs.	4375	4375	3634	3634

Notes: OLS regression. The dependent variable is indicated in the column headings. 'Correct answer' is a dummy capturing whether the respondent recalled the wage offered correctly. 'Absolute mistake' is the absolute difference between the wage recalled by the respondent and the wage actually offered. The number of observation changes because some individuals report that they do not remember the wage offered. These individuals are included in the regressions reported in columns (1) and (2), but not in the regressions reported in columns (3) and (4). Columns (2) and (4) include a control for whether the respondent has applied for the experiment's job. Robust standard errors reported in parenthesis. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Sample used: second phone call sample.

Table A.39: Beliefs about labour-market prospects

	Weeks unemployment	Wage (Ethiopian Birr)
	(1)	(2)
Incentive	-.043	102.288
	(0.483)	(173.889)
High Wage	-.642	445.66
	(0.533)	(128.709)
Control group mean	8.69	5192.291
Incentive = Wage (p)	0.235	0.051
Obs.	3849	3817

Notes: OLS regression. 'Weeks unemployment' captures the number of weeks that the respondent expect she or he would need in order to be offered a job they would be willing to work at. 'Wage' captures the wage that the respondent expects this job will pay. Beliefs about the wage are elicited through the method of Attanasio and Kaufmann (2009), as explained in footnote 33. For both questions the respondent was asked to consider an hypothetical job search spell starting on the day following their interview. Thus, the answers to these questions do not refer to the experiment's job. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.40: Beliefs about the probability of getting the experiment's job

	Subjective forecast		
	(1)	(2)	(3)
Incentive	1.13	.619	.105
	(1.119)	(1.095)	(1.071)
High Wage	1.461	1.453	.648
	(1.112)	(1.090)	(1.070)
Forecasts implying certainty	included	excluded	excluded
Truncation	no	no	yes
Control group mean	57.065	53.083	48.544
Incentive = Wage (p)	0.769	0.450	0.617
Obs.	4325	3893	3446

Notes: OLS regression. The dependent variable is the subject's subjective forecast of the probability of being offered the experiment's job. We elicit this forecast retrospectively, by asking the following question in the second phone call: "How confident were you of getting an offer for this position at the time when you decided whether to apply or not? In order to quantify this, you can think of applying to 100 positions like this one. How many offers would you get?". In the first column, we report the raw data. In the second and third column, we drop forecasts that imply certainty, that is, forecasts of 0 or 100. In the third column, we truncate the variable at the 95th percentile. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.41: Beliefs about the attributes of the job

	Holidays	Overtime	Satisfaction	Autonomy	Career	Opportunities	New Skills
	(1)	(2)	(3)	(4)	(5)	(6)	(8)
Incentive	.023	.019	.007	-.012	.035	.006	-.005
	(0.012)	(0.017)	(0.010)	(0.017)	(0.013)	(0.010)	(0.004)
High Wage	.016	.045	.028	-.007	.043	.003	.001
	(0.011)	(0.017)	(0.009)	(0.017)	(0.012)	(0.010)	(0.004)
Control group mean	0.127	0.411	0.904	0.486	0.810	0.908	0.986
Incentive = Wage (p)	0.514	0.125	0.019	0.726	0.495	0.766	0.118
Obs.	4366	4362	4364	4361	4363	4364	4368

Notes: OLS regression. The dependent variable is indicated in the column headings. 'Holiday' is a dummy variable capturing whether the respondent believes the job has more than four days of holiday per month. 'Overtime' is a dummy variable capturing whether the respondent believes the job will require work in the evenings. 'Satisfaction' is a dummy variable capturing whether the respondent believes the job will be satisfying. 'Autonomy' is a dummy variable capturing whether the respondent believes he or she will have freedom to organise their own schedule at work. 'Career' is a dummy variable capturing whether the respondent believes the experience in this job will help them find other jobs in the future. 'Opportunity' is a dummy variable capturing whether the respondent believes there will be further work opportunities with the employer. 'New Skills' is a dummy variable capturing whether the respondent believes they will learn new skills in this job. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: second phone call sample.

Table A.42: Survey experiment on job-attribute beliefs: Balance

	Mean		StDev	N	Balance test (p)
	Control	Incentive			
	(1)	(2)	(3)	(4)	(5)
Female	0.49	0.47	0.50	724	0.66
Age	23.95	24.13	2.21	724	0.28
Born in Addis Ababa	0.20	0.17	0.40	724	0.21
Wage work experience (dummy)	0.40	0.43	0.49	724	0.45
Wage work experience (months)	20.98	20.90	23.04	235	0.98
Currently unemployed	0.77	0.82	0.42	724	0.08
Overall balance					0.22

Notes: In this Table, we present summary and balance statistics for the sample of individuals that participated in the new survey fielded in 2019/2020. We present summary statistics in columns 1-4. In column 5, we report the p-value of a test of covariate balance. We first report balance tests for single covariates and then, in the last row, report a joint test of orthogonality (following the recent literature, e.g. McKenzie (2017b)). To perform the joint test of orthogonality we regress the treatment variable on all covariates and we then test the joint hypothesis that all covariates have a zero coefficient. Sample used: 2020 survey sample.

Table A.43: Survey experiment on job-attribute beliefs: Results

	Expectations					
	Holidays	Overtime	Satisfaction	Autonomy	Career	Enjoyable
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.019	.011	-.017	-.017	.033	-.003
	(.037)	(.034)	(.037)	(.029)	(.034)	(.037)
Control mean	0.461	0.279	0.459	0.202	0.691	0.544
Obs.	724	724	724	724	724	724

Notes: OLS regression. The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. 'Holiday' is a dummy variable capturing whether the respondent believes the job has more than four days of holiday per month. 'Overtime' is a dummy variable capturing whether the respondent believes the job will require work in the evenings. 'Satisfaction' is a dummy variable capturing whether the respondent believes the job will be satisfying. 'Autonomy' is a dummy variable capturing whether the respondent believes he or she will have freedom to organise their own schedule at work. 'Career' is a dummy variable capturing whether the respondent believes the experience in this job will help them find other jobs in the future. 'Enjoyable' is a dummy variable capturing whether the respondent believes the work environment will be pleasant. Robust standard errors reported in parenthesis. Sample used: 2020 survey sample.

Table A.44: Impacts on individual psychological traits

	Conscientiousness	Neuroticism	Openness	Extraversion	Agreeableness	Grit	Locus of control	Core self-evaluaton	Self-esteem
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Incentive	.005	.02	.02	-.011	.041	-.033	-.114	-.152	-.04
	(0.026)	(0.033)	(0.022)	(0.025)	(0.024)	(0.025)	(0.517)	(0.185)	(0.028)
High Wage	0.032	-0.032	0.030	0.018	0.041	0.029	-0.816	-0.010	0.019
	(0.024)	(0.031)	(0.021)	(0.024)	(0.023)	(0.024)	(0.490)	(0.183)	(0.026)
Control Mean	4.539	1.762	4.115	3.467	4.244	3.862	35.309	23.431	4.121
St. dev.	.488	.62	.435	.464	.481	.475	9.817	3.545	.52
Incentive = Wage (p)	0.234	0.076	0.573	0.193	0.995	0.006	0.120	0.396	0.015
Obs.	2385	2385	2385	2385	2385	2373	2378	2377	2363

Notes: OLS regressions. The dependent variable is reported in the column headings. The first five columns report impacts on each of the big five personality traits (John and Srivastava, 1999). Columns (6)-(9) report impacts on grit (Duckworth et al., 2007), core self evaluation (Gardner and Pierce, 2010), locus of control scale (Lefcourt, 1991) and self-esteem (Rosenberg, 1986). The second to last row reports the p-value of a test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.45: Impacts on specific work experience

	Routine	Physical	Managerial	Problem-solving	Math	Reading	Client
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentive	-0.172	0.599	-0.816	-1.320	-1.157	0.450	0.041
	(1.864)	(1.121)	(1.172)	(1.731)	(1.584)	(1.829)	(0.026)
High Wage	-0.715	2.870	-0.159	-0.939	-0.564	-0.096	0.027
	(1.731)	(2.143)	(1.179)	(1.724)	(1.577)	(1.660)	(0.025)
Control Mean	13.029	6.480	8.566	11.328	8.534	10.904	0.347
St. dev.	36.669	19.374	24.005	37.380	34.780	35.069	0.477
Incentive = Wage (p)	0.707	0.291	0.484	0.748	0.563	0.706	0.557
Obs.	2311	2310	2311	2311	2311	2311	2310

Notes: OLS regressions. The dependent variable is reported in the column headings. All dependent variables refer to specific dimensions of job experience, measured using the questionnaire developed by Autor and Handel (2013). The variables reported in columns (1)-(6) measure the number of months of experience in a job where the task indicated in the variable name was regularly performed. 'Routine' refers to short, repetitive tasks; 'managerial' refers to tasks such as managing and supervising other workers; 'physical' refers to physical tasks such as standing, handling objects, operating machinery or vehicles, or making or fixing things with one's hands; 'problem-solving' refers to tasks where the worker is faced with a new or difficult situation where they have to think for at least 30 minutes to find a good solution; 'math' refers to advanced mathematics such as algebra, geometry, trigonometry, probability, or calculus; 'reading' refers to tasks that require reading documents that are longer than ten pages. Column (7), 'client-service', is a dummy variable that takes a value of one if the respondent has substantial experience interacting face-to-face with clients, customers, suppliers or contractors. Robust standard errors reported in parenthesis. Sample used: all applicants.

Table A.46: Identification (noisy-ability case)

Structural parameters (13)		Moments (14/16)
Quality and costs:	$\Leftrightarrow$	$\text{E}[T \mid apply, B = b_z, \texttt{control}]$,
μ_{T_l}, σ_{T_l}, μ_{C_l},		$\text{SD}[T \mid apply, B = b_z, \texttt{control}]$,
μ_{T_h}, σ_{T_h}, μ_{C_h}		$\Pr[apply \mid B = b_z, \texttt{control}]$
		for $z \in \{l, h\}$
Shocks and st. dev. of costs:	$\Leftrightarrow$	$\Delta\text{Applications}[B = b_z, \texttt{incentive}]$
σ_{C_l}, σ_{C_h}, τ, τ^w		$\Delta\text{Applications}[B = b_z, \texttt{wage}]$
		for $z \in \{l, h\}$
Covariance and selectivity:	$\Leftrightarrow$	$\Delta\text{ApplicantAbility}[B = b_z, \texttt{incentive}]$,
σ_{TC_l}, σ_{TC_h}, a		$\Delta\text{ApplicantAbility}[B = b_z, \texttt{wage}]$
		$\text{E}[\Pr[T > a \mid B = b_z, C = c]]$
		for $z \in \{l, h\}$

Table A.47: Identification (noisy-selection case)

Structural parameters (14)		Moments (14/16)
Quality and costs:	$\Leftrightarrow$	$\mathrm{E}[T\|apply, B = b_z, \texttt{control}]$,
μ_{T_l}, σ_{T_l}, μ_{C_l},		$\mathrm{SD}[T\|apply, B = b_z, \texttt{control}]$,
μ_{T_h}, σ_{T_h}, μ_{C_h}		$\Pr[apply\|B = b_z, \texttt{control}]$
		for $z \in \{l, h\}$
Shocks and st. dev. of costs:	$\Leftrightarrow$	ΔApplications$[B = b_z, \texttt{incentive}]$
σ_{C_l}, σ_{C_h}, τ, τ^w		ΔApplications$[B = b_z, \texttt{wage}]$
		for $z \in \{l, h\}$
Covariance and selectivity:	$\Leftrightarrow$	ΔApplicantAbility$[B = b_z, \texttt{incentive}]$,
σ_{TC_l}, σ_{TC_h}, μ_a, σ_a		ΔApplicantAbility$[B = b_z, \texttt{wage}]$
		$\mathrm{E}[\Pr[T > a]]$
		for $z \in \{l, h\}$

Table A.48: Additional parameter estimates

	(1)	(2)	(3)	(4)
	Low B			
μ_T	45.605	45.697	45.165	45.357
	(1.11)	(0.97)	(2.76)	(3.15)
σ_T	13.785	13.818	13.574	13.667
	(0.76)	(0.73)	(1.58)	(1.75)
	High B			
μ_T	46.365	46.457	47.231	46.905
	(1.03)	(1.03)	(3.64)	(3.59)
σ_T	14.293	14.348	14.701	14.541
	(0.87)	(0.87)	(2.26)	(2.21)
Information	Noisy ability		Noisy selectivity	
Moments	14	16	14	16

Notes: The table shows the additional parameter on the marginal distribution of ability that we did not report in Table 4. We report both estimates for the noisy-ability case (columns 1 and 2) and the noisy-selection case (columns 3 and 4). Estimation is based on minimum distance estimation. Column (1) and (3) use 14 moments (reported in Table A.49 and Table A.51). Column (2) and (4) use 16 moments (reported in Table A.50 and Table A.52). Standard errors obtained through a bootstrap of the structural estimation reported in parenthesis. The bootstrap includes the estimation of B and the demediation procedure.

Table A.49: Fit between empirical and simulated moments (noisy-ability case, core moments)

Moment	Empirical	Simulated
Low B		
$\Pr[apply \mid B = b_l, \texttt{control}]$	47.648	47.651
$\mathrm{E}[T \mid apply, B = b_l, \texttt{control}]$	38.225	38.234
$\mathrm{SD}[T \mid apply, B = b_l, \texttt{control}]$	11.811	11.812
ΔApplications$[B = b_l, \texttt{incentive}]$	10.952	11.165
ΔApplications$[B = b_l, \texttt{wage}]$	14.286	13.713
ΔAbility$[B = b_l, \texttt{incentive}]$	1.973	1.536
ΔAbility$[B = b_l, \texttt{wage}]$	1.462	1.872
High B		
$\Pr[apply \mid B = b_h, \texttt{control}]$	49.667	49.662
$\mathrm{E}[T \mid apply, B = b_h, \texttt{control}]$	39.812	39.803
$\mathrm{SD}[T \mid apply, B = b_h, \texttt{control}]$	12.715	12.715
ΔApplications$[B = b_h, \texttt{incentive}]$	8.337	8.749
ΔApplications$[B = b_h, \texttt{wage}]$	12.033	11.291
ΔAbility$[B = b_h, \texttt{incentive}]$	2.332	1.107
ΔAbility$[B = b_h, \texttt{wage}]$	0.581	1.418

Notes: The table shows the empirical and simulated moments for the structural estimates reported in column (1) of Table 4 (noisy-ability case, core moments). $\Pr[apply \mid B = b_l, \texttt{control}$ is the application rate for low-B jobseekers in the control group. $\mathrm{E}[T \mid apply, B = b_l, \texttt{control}]$ is the average ability among control applicants from the low-B group. $\mathrm{SD}[T \mid apply, B = b_l, \texttt{control}]$ is the standard deviation of ability among control applicants from the low-B group. ΔApplications$[B = b_l, \texttt{incentive}]$ is the change in application rates generated by the incentive intervention among low-B jobseekers. ΔApplications$[B = b_l, \texttt{wage}]$ is the change in application rates generated by the high wage intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{incentive}]$ is the change in average applicant ability generated by the incentive intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{wage}]$ is the change in average applicant ability generated by the high wage intervention among low-B jobseekers. Moments for high-B jobseekers are defined in a similar way. To generate the two groups, we first drop observations with a negative estimated value of B. We then split the remaining observations at the median value of B.

Table A.50: Fit between empirical and simulated moments (noisy-ability case, core moments + beliefs)

Moment	Empirical	Simulated
Low B		
Pr[$apply$|$B = b_l$, $\texttt{control}$]	47.648	47.626
E[T|$apply, B = b_l$, $\texttt{control}$]	38.225	38.270
SD[T|$apply, B = b_l$, $\texttt{control}$]	11.811	11.820
ΔApplications[$B = b_l$, $\texttt{incentive}$]	10.952	11.117
ΔApplications[$B = b_l$, $\texttt{wage}$]	14.286	13.785
ΔAbility[$B = b_l$, $\texttt{incentive}$]	1.973	1.540
ΔAbility[$B = b_l$, $\texttt{wage}$]	1.462	1.895
E[$Pr[T > a]$|$C = c, B = b_l$]	0.482	0.464
High B		
Pr[$apply$|$B = b_h$, $\texttt{control}$]	49.667	49.689
E[T|$apply, B = b_h$, $\texttt{control}$]	39.812	39.762
SD[T|$apply, B = b_h$, $\texttt{control}$]	12.715	12.708
ΔApplications[$B = b_h$, $\texttt{incentive}$]	8.337	8.811
ΔApplications[$B = b_h$, $\texttt{wage}$]	12.033	11.119
ΔAbility[$B = b_h$, $\texttt{incentive}$]	2.332	1.138
ΔAbility[$B = b_h$, $\texttt{wage}$]	0.581	1.426
E[$Pr[T > a]$|$C = c, B = b_h$]	0.468	0.487

Notes: The table shows the empirical and simulated moments for the structural estimates reported in column (2) of Table 4 (noisy-ability case, core moments + beliefs). Pr[$apply|B = b_l$, $\texttt{control}$ is the application rate for low-B jobseekers in the control group. E[$T|apply, B = b_l$, $\texttt{control}$] is the average ability among control applicants from the low-B group. SD[$T|apply, B = b_l$, $\texttt{control}$] is the standard deviation of ability among control applicants from the low-B group. ΔApplications[$B = b_l$, $\texttt{incentive}$] is the change in application rates generated by the incentive intervention among low-B jobseekers. ΔApplications[$B = b_l$, $\texttt{wage}$] is the change in application rates generated by the high wage intervention among low-B jobseekers. ΔAbility[$B = b_l$, $\texttt{incentive}$] is the change in average applicant ability generated by the incentive intervention among low-B jobseekers. ΔAbility[$B = b_l$, $\texttt{wage}$] is the change in average applicant ability generated by the high wage intervention among low-B jobseekers. $E[Pr[T > a]|C = c, B = b_l]$ is the average forecast of the probability of being offered the job among low-B jobseekers. Moments for high-B jobseekers are defined in a similar way. To generate the two groups, we first drop observations with a negative estimated value of B. We then split the remaining observations at the median value of B.

Table A.51: Fit between empirical and simulated moments (noisy-selection case, core moments)

Moment	Empirical	Simulated
Low B		
Pr$[apply\|B = b_l, \texttt{control}]$	47.648	47.646
E$[T\|apply, B = b_l, \texttt{control}]$	38.225	38.225
SD$[T\|apply, B = b_l, \texttt{control}]$	11.811	11.811
ΔApplications$[B = b_l, \texttt{incentive}]$	10.952	11.185
ΔApplications$[B = b_l, \texttt{wage}]$	14.286	13.930
ΔAbility$[B = b_l, \texttt{incentive}]$	1.973	1.448
ΔAbility$[B = b_l, \texttt{wage}]$	1.462	1.790
High B		
Pr$[apply\|B = b_h, \texttt{control}]$	49.667	49.667
E$[T\|apply, B = b_h, \texttt{control}]$	39.812	39.813
SD$[T\|apply, B = b_h, \texttt{control}]$	12.715	12.715
ΔApplications$[B = b_h, \texttt{incentive}]$	8.337	8.754
ΔApplications$[B = b_h, \texttt{wage}]$	12.033	10.914
ΔAbility$[B = b_h, \texttt{incentive}]$	2.332	1.252
ΔAbility$[B = b_h, \texttt{wage}]$	0.581	1.553

Notes: The table shows the empirical and simulated moments for the structural estimates reported in column (3) of Table 4 (noisy-selection case, core moments). Pr$[apply|B = b_l, \texttt{control}$ is the application rate for low-B jobseekers in the control group. E$[T|apply, B = b_l, \texttt{control}]$ is the average ability among control applicants from the low-B group. SD$[T|apply, B = b_l, \texttt{control}]$ is the standard deviation of ability among control applicants from the low-B group. ΔApplications$[B = b_l, \texttt{incentive}]$ is the change in application rates generated by the incentive intervention among low-B jobseekers. ΔApplications$[B = b_l, \texttt{wage}]$ is the change in application rates generated by the high wage intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{incentive}]$ is the change in average applicant ability generated by the incentive intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{wage}]$ is the change in average applicant ability generated by the high wage intervention among low-B jobseekers. Moments for high-B jobseekers are defined in a similar way. To generate the two groups, we first drop observations with a negative estimated value of B. We then split the remaining observations at the median value of B.

Table A.52: Fit between empirical and simulated moments (noisy-selection case, core moments + beliefs)

Moment	Empirical	Simulated
Low B		
Pr$[apply \| B = b_l, \texttt{control}]$	47.648	47.647
E$[T \| apply, B = b_l, \texttt{control}]$	38.225	38.225
SD$[T \| apply, B = b_l, \texttt{control}]$	11.811	11.811
ΔApplications$[B = b_l, \texttt{incentive}]$	10.952	11.132
ΔApplications$[B = b_l, \texttt{wage}]$	14.286	13.881
ΔAbility$[B = b_l, \texttt{incentive}]$	1.973	1.482
ΔAbility$[B = b_l, \texttt{wage}]$	1.462	1.834
E$[Pr[T > a]]$	0.482	0.478
High B		
Pr$[apply \| B = b_h, \texttt{control}]$	49.667	49.666
E$[T \| apply, B = b_h, \texttt{control}]$	39.812	39.812
SD$[T \| apply, B = b_h, \texttt{control}]$	12.715	12.715
ΔApplications$[B = b_h, \texttt{incentive}]$	8.337	8.812
ΔApplications$[B = b_h, \texttt{wage}]$	12.033	10.999
ΔAbility$[B = b_h, \texttt{incentive}]$	2.332	1.205
ΔAbility$[B = b_h, \texttt{wage}]$	0.581	1.496
E$[Pr[T > a]]$	0.468	0.478

Notes: The table shows the empirical and simulated moments for the structural estimates reported in column (4) of Table 4 (noisy-selection case, core moments + beliefs). Pr$[apply|B = b_l, \texttt{control}$ is the application rate for low-B jobseekers in the control group. E$[T|apply, B = b_l, \texttt{control}]$ is the average ability among control applicants from the low-B group. SD$[T|apply, B = b_l, \texttt{control}]$ is the standard deviation of ability among control applicants from the low-B group. ΔApplications$[B = b_l, \texttt{incentive}]$ is the change in application rates generated by the incentive intervention among low-B jobseekers. ΔApplications$[B = b_l, \texttt{wage}]$ is the change in application rates generated by the high wage intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{incentive}]$ is the change in average applicant ability generated by the incentive intervention among low-B jobseekers. ΔAbility$[B = b_l, \texttt{wage}]$ is the change in average applicant ability generated by the high wage intervention among low-B jobseekers. $E[Pr[T > a]|B = b_l]$ is the average forecast of the probability of being offered the job among low-B jobseekers. Moments for high-B jobseekers are defined in a similar way. To generate the two groups, we first drop observations with a negative estimated value of B. We then split the remaining observations at the median value of B.

Table A.53: Elasticity of simulated moments
(noisy-ability case, core moments)

	μ_T	σ_T	μ_C	σ_C	σ_{TC}	a	τ	τ^w
E$[T\|apply, B = b_h, \texttt{control}]$	2.487	0.163	0.317	0.219	0.226	1.339	0.000	0.000
SD$[T\|apply, B = b_h, \texttt{control}]$	0.472	1.321	0.110	0.275	0.283	0.440	0.000	0.000
Pr$[apply\|B = b_h, \texttt{control}]$	8.246	1.005	1.929	0.346	0.364	8.048	0.000	0.000
ΔApplications$[B = b_h, \texttt{incentive}]$	3.095	2.190	0.672	5.204	2.726	6.111	1.021	0.000
ΔApplications$[B = b_h, \texttt{wage}]$	15.570	1.612	2.595	6.191	3.233	18.728	0.000	1.249
ΔAbility$[B = b_h, \texttt{incentive}]$	0.506	2.466	0.027	6.124	3.802	2.953	0.993	0.000
ΔAbility$[B = b_h, \texttt{wage}]$	12.593	1.869	1.890	7.037	4.287	15.682	0.000	1.213

Notes: The table reports the moment elasticity for the high-B group and the noisy-ability case, estimated using fourteen moments. The corresponding parameter estimates are reported in column (1) of Table 4. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the structural estimates of the parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments.

Table A.54: Elasticity of simulated moments
(noisy-ability case, core moments + beliefs)

	μ_T	σ_T	μ_C	σ_C	σ_{TC}	a	τ	τ^w
E$[T \mid apply, B = b_h, \texttt{control}]$	2.907	0.030	0.546	0.184	0.181	1.798	0.000	0.000
SD$[T \mid apply, B = b_h, \texttt{control}]$	0.637	1.283	0.189	0.275	0.283	0.598	0.000	0.000
Pr$[apply \mid B = b_h, \texttt{control}]$	10.652	0.177	3.256	0.091	0.083	10.529	0.000	0.000
ΔApplications$[B = b_h, \texttt{incentive}]$	5.070	3.353	0.463	6.325	3.655	0.598	0.974	0.000
ΔApplications$[B = b_h, \texttt{wage}]$	2.725	3.831	1.547	7.240	4.317	10.806	0.000	1.106
ΔAbility$[B = b_h, \texttt{incentive}]$	7.858	3.322	1.626	7.120	4.615	3.674	0.940	0.000
ΔAbility$[B = b_h, \texttt{wage}]$	0.196	3.773	0.407	7.980	5.245	6.781	0.000	1.073
$E[Pr[T > a] \mid C = c, B = b_h]$	1.326	0.013	0.000	0.000	0.000	2.790	0.000	0.000

Notes: The table reports the moment elasticity for the high-B group and the noisy-ability case, estimated using sixteen moments. The corresponding parameter estimates are reported in column (2) of Table 4. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the structural estimates of the parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments.

Table A.55: Elasticity of simulated moments
(noisy-selection case, core moments)

	μ_T	σ_T	μ_C	σ_C	σ_{TC}	μ_a	σ_a	τ	τ^w
E$[T \mid apply, B = b_h, \texttt{control}]$	1.186	0.000	0.224	0.186	0.188	0.045	0.048	0.000	0.000
SD$[T \mid apply, B = b_h, \texttt{control}]$	0.000	1.337	0.087	0.330	0.338	0.016	0.024	0.000	0.000
Pr$[apply \mid B = b_h, \texttt{control}]$	0.004	0.000	1.202	0.008	0.000	0.252	0.250	0.000	0.000
ΔApplications$[B = b_h, \texttt{incentive}]$	0.001	0.000	0.141	0.978	0.000	0.033	0.030	0.984	0.000
ΔApplications$[B = b_h, \texttt{wage}]$	0.000	0.000	0.183	0.971	0.000	0.165	0.165	0.000	0.971
ΔAbility$[B = b_h, \texttt{incentive}]$	0.000	0.000	0.559	1.932	1.006	0.112	0.120	0.958	0.000
ΔAbility$[B = b_h, \texttt{wage}]$	0.000	0.000	0.573	1.926	0.998	0.077	0.077	0.000	0.947

Notes: The table reports the moment elasticity for the high-B group and the noisy-selection case, estimated using fourteen moments. The corresponding parameter estimates are reported in column (3) of Table 4. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the structural estimates of the parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments.

Table A.56: Elasticity of simulated moments
(noisy-selection case, core moments + beliefs)

	μ_T	σ_T	μ_C	σ_C	σ_{TC}	μ_a	σ_a	τ	τ^w
E$[T \vert apply, B = b_h, \texttt{control}]$	1.181	0.003	0.153	0.178	0.178	0.008	0.008	0.000	0.000
SD$[T \vert apply, B = b_h, \texttt{control}]$	0.000	1.306	0.055	0.299	0.315	0.008	0.000	0.000	0.000
Pr$[apply \vert B = b_h, \texttt{control}]$	0.002	0.000	0.874	0.006	0.000	0.042	0.040	0.000	0.000
ΔApplications$[B = b_h, \texttt{incentive}]$	0.000	0.000	0.107	0.976	0.001	0.006	0.005	0.985	0.000
ΔApplications$[B = b_h, \texttt{wage}]$	0.000	0.000	0.136	0.973	0.000	0.045	0.036	0.000	0.973
ΔAbility$[B = b_h, \texttt{incentive}]$	0.000	0.000	0.407	1.933	1.004	0.017	0.017	0.954	0.000
ΔAbility$[B = b_h, \texttt{wage}]$	0.007	0.000	0.421	1.919	1.003	0.027	0.027	0.000	0.949
$E[Pr[T > a]]$	0.001	0.000	0.000	0.000	0.000	0.049	0.046	0.000	0.000

Notes: The table reports the moment elasticity for the high-B group and the noisy-selection case, estimated using sixteen moments. he corresponding parameter estimates are reported in column (4) of Table 4. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the structural estimates of the parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments.

Table A.57: Robustness of parameter estimates (noisy-ability case)

	(1)	(2)	(3)	(4)	(5)	(6)
			Low B			Low B
μ_C	112.270	138.960	136.100	136.260	115.130	153.850
σ_C	190.750	193.550	192.980	192.750	244.900	215.800
ρ	0.733	0.638	0.640	0.640	0.673	0.504
			High B			Medium B
μ_C	206.860	215.600	217.510	217.350	133.090	216.360
σ_C	213.610	244.690	244.930	245.220	203.250	235.530
ρ	0.587	0.572	0.572	0.571	0.506	0.649
						High B
μ_C						274.760
σ_C						235.130
ρ						0.580
a	2.479	50.205	50.431	50.429	50.239	48.115
		50.538				
τ	10.889	19.112	19.092	18.971	39.747	19.856
				19.191		
τ^w	34.077	55.449	56.187	55.677	127.140	28.454
			55.225			
Goodness of fit	.34147	2.2612	2.2618	2.2626	4.8354	3.6019

Notes: The table shows parameter estimates for the noisy-ability model. Estimation is based on minimum distance estimation. The model in column (1) uses moments based on the cognitive ability score (as opposed to the Raven test score). The model in columns (2)-(4) use the 14 moments reported in Table A.49. These models let, in turn, a, τ^w, and τ differ by B group. For each model, we first report the value of the parameter for the low-B group and, in the row below, we report the value for the high-B group. The model in column (5) uses moments obtained by predicting B using an OLS model instead of the post-LASSO estimator. The model in column (6) allows for three types of B. Costs are expressed in Ethiopian Birr.

Table A.58: Parameter estimates: heterogeneity by age

	Old	Young
μ_T	45.475	45.653
σ_T	13.136	14.338
μ_C	200.670	262.760
σ_C	223.350	241.070
ρ	0.578	0.580
a	46.315	
τ	23.487	
τ^w	54.258	
Goodness of fit	4.7357	

Notes: Estimates from classical minimum distance estimator. Noisy-ability case. Empirical moments obtained by splitting the sample by age. Empirical and simulated moments reported in Table A.60. Costs are expressed in Ethiopian Birr.

Table A.59: Parameter estimates: heterogeneity by gender

	Men	Women
μ_T	46.508	44.355
σ_T	13.628	14.911
μ_C	241.340	211.500
σ_C	225.580	219.840
ρ	0.616	0.598
a	46.923	
τ	17.424	
τ^w	44.299	
Goodness of fit	18.535	

Notes: Estimates from classical minimum distance estimator. Noisy-ability case. Empirical moments obtained by splitting the sample by age. Empirical and simulated moments reported in Table A.61. Costs are expressed in Ethiopian Birr.

Table A.60: Fit between empirical and simulated moments: heterogeneity by age (noisy-ability case, core moments + beliefs)

Moment	Empirical	Simulated
	Old	
$\Pr[apply \| \texttt{old}, \texttt{control}]$	48.611	48.665
$\mathrm{E}[T \| apply, \texttt{old}, \texttt{control}]$	39.132	39.253
$\mathrm{SD}[T \| apply, \texttt{old}, \texttt{control}]$	11.656	11.637
ΔApplications$[\texttt{old}, \texttt{incentive}]$	8.213	8.910
ΔApplications$[\texttt{old}, \texttt{wage}]$	13.606	11.031
ΔAbility$[\texttt{old}, \texttt{incentive}]$	0.618	1.055
ΔAbility$[\texttt{old}, \texttt{wage}]$	0.265	1.298
$\mathrm{E}[Pr[T > a] \| C = c, \texttt{old}]$	0.474	0.475
	Young	
$\Pr[apply \| \texttt{young}, \texttt{control}]$	48.621	48.580
$\mathrm{E}[T \| apply, \texttt{young}, \texttt{control}]$	38.929	38.825
$\mathrm{SD}[T \| apply, \texttt{young}, \texttt{control}]$	12.666	12.688
ΔApplications$[\texttt{young}, \texttt{incentive}]$	10.645	10.718
ΔApplications$[\texttt{young}, \texttt{wage}]$	12.919	13.903
ΔAbility$[\texttt{young}, \texttt{incentive}]$	2.974	1.383
ΔAbility$[\texttt{young}, \texttt{wage}]$	1.540	1.778
$\mathrm{E}[Pr[T > a] \| C = c, \texttt{young}]$	0.475	0.482

Notes: The table shows the empirical and simulated moments for the structural estimates reported in Table A.58.

Table A.61: Fit between empirical and simulated moments: heterogeneity by gender (noisy-ability case, core moments + beliefs)

Moment	Empirical	Simulated
Men		
Pr[$apply$|$\texttt{male}, \texttt{control}$]	48.028	48.083
E[T|$apply, \texttt{male}, \texttt{control}$]	39.656	39.554
SD[T|$apply, \texttt{male}, \texttt{control}$]	11.992	11.839
ΔApplications[$\texttt{male}, \texttt{incentive}$]	10.677	9.498
ΔApplications[$\texttt{male}, \texttt{wage}$]	14.409	13.749
ΔAbility[$\texttt{male}, \texttt{incentive}$]	0.812	1.246
ΔAbility[$\texttt{male}, \texttt{wage}$]	0.668	1.780
E[$Pr[T > a]$|$C = c, \texttt{male}$]	0.475	0.488
Women		
Pr[$apply$|$\texttt{female}, \texttt{control}$]	50.435	50.335
E[T|$apply, \texttt{female}, \texttt{control}$]	37.087	37.286
SD[T|$apply, \texttt{female}, \texttt{control}$]	12.945	13.108
ΔApplications[$\texttt{female}, \texttt{incentive}$]	6.769	9.981
ΔApplications[$\texttt{female}, \texttt{wage}$]	13.872	13.044
ΔAbility[$\texttt{female}, \texttt{incentive}$]	5.898	1.368
ΔAbility[$\texttt{female}, \texttt{wage}$]	2.035	1.773
E[$Pr[T > a]$|$C = c, \texttt{female}$]	0.473	0.432

Notes: The table shows the empirical and simulated moments for the structural estimates reported in Table A.59.

A.3 Measures of ability, preferences and sophistication

A.3.1 Cognitive ability

We administer a Raven test and a Stroop test. The Raven test consists of 60 questions (Raven, 2000). Participants are given basic instructions about the test from an instructor and then have to complete the test in 60 minutes. To measure performance on this test, we use the number of correct answers.

We administer the Stroop test proposed by Mani et al. (2013). In this test, the instructor shows a string of digits and then test-taker has to report the number of digits shown. For example, if the string is '44', the correct answer is 'two'. Individuals are shown 75 strings in total. There are two measures of performance: the number of mistakes and (ii) the time taken to complete all strings (which is measured by the instructor using a stopwatch).

A.3.2 Non-cognitive ability

Our main measures of non-cognitive ability are derived from two standard scales: the big five inventory (BFI-44) and the 12-item grit scale (John and Srivastava, 1999; Duckworth et al., 2007). Further, we administer the 12-item core self evaluation scale (Gardner and Pierce, 2010), a 16-item locus of control scale (Lefcourt, 1991), Rosenberg?s 10-item self-esteem scale (Rosenberg, 1986). Participants are told that their answers to these psychometric questions are not going to be used to select the workers for the position. We included this feature to maximise truthful reporting.

A.3.3 Time preferences

We measure time preferences over the allocation of effort using a design proposed by Augenblick et al. (2015). We administer this task after the main measures of ability are collected. Applicants are informed that with a certain probability they will be invited to complete a small job, for which they will receive a financial remuneration. This job consists of transcribing 60 pages of text. The job has to be completed in two separate sessions, one week apart from each other. Participants have to transcribe at least five pages per session, but are free to allocate the remaining 50 pages across the two sessions. They are informed that this job is unrelated to the main position and that the effort allocation decisions is not going to be used in the selection process for the main position.

We ask individuals to make allocation decisions for ten different scenarios, one of which will be randomly drawn and implemented. In the first five scenarios, the near work session is on the day following the allocation decision and the late work session is seven days after that. In the last five scenarios, the near work session is two weeks after the allocation decision, and the late work session is seven days after that.[51] Across scenarios, we vary the relative cost of allocating work to early and late sessions. Pages allocated to the near work session always have four sentences. Pages allocated to the late work sessions have $x =$ 6,5,4,3 or 2 sentences, depending on the scenario. $R = \frac{4}{x}$ is thus the rate of exchange of effort between the late and the early work session.

Consider an individual with beta-delta preferences and a cost of effort function given by $(e + \omega)^\gamma$ (where e is the effort chosen and ω is background effort, in our case 5 pages). Augenblick et al. (2015) show that, for each scenario d, the allocation of effort between the near work session at time t and the late work session at time $t + k$ is given by:

$$\log \frac{e_{d,t} + \omega}{e_{d,t+k} + \omega} = \frac{\log(\beta)}{\gamma - 1} \text{Early}_d + \frac{\log(\delta)}{\gamma - 1} \text{k}_d + \frac{1}{\gamma - 1} \log(\text{R})_d \tag{A.1}$$

where Early_d is a dummy for scenarios where the near date is on the following day. We estimate equation (A.1) for each experimental group using a two-limit tobit estimator and obtain estimates of β, γ and δ through non-linear combination of the coefficients. We obtain standard errors and test hypothesis about the equality of the coefficients using the delta method. We also obtain individual estimates of each parameter by estimating model (A.1) for each individual. As these estimates are less stable, we windsorize the estimates of β_i and classify as present biased any individual with $\beta < .99$.

A.3.4 Risk and social preferences

We measure risk preferences using the following questions adapted from the Global Preferences Survey (Falk et al., 2016):

1. How do you see yourself: are you a person who is generally willing to take risks, or do you try to avoid taking risks? (On a scale from 0 to 10).

2. Please imagine the following situation: You can choose between a sure payment and a lottery. The lottery gives you a 50 percent chance of receiving 500 Birr. With

[51] Differences in allocation decisions in near and far time horizons enable us to identify present bias. This is a common strategy in the literature on time preferences (Dohmen et al., 2012). Augenblick et al. (2015), on the other hand, elicit allocation decisions for a single time horizon and enable subjects to revise this decision just before the start of the first work session. This feature of the design of Augenblick et al. (2015) would have been difficult to replicate in our setting.

an equally high chance you receive nothing. Now imagine you had to choose between the lottery and a sure payment. We will present to you 2 different situations. The lottery is the same in all situations. The sure payment is different in every situation.[52]

3. Please imagine the following situation: you have won a prize in a contest. Now you can choose between two different payment methods, either a lottery or a sure payment. If you choose the lottery there is a 50 percent chance that you receive 1700 Birr and an equally high chance that you receive nothing. Please consider: what would the sure payment need to be in order for you to prefer the sure payment over playing the lottery?

We measure social preferences using the following questions adapted from the Global Preferences Survey (Falk et al., 2016):

1. How would you assess your willingness to share with others without expecting anything in return, for example your willingness to give to charity? (On a scale from 0 to 10).

2. Imagine the following situation: Today you unexpectedly received 1700 Birr. How much of this amount would you donate to charity?

3. How well does the following statement describe you as a person? I do not understand why people spend a lifetime fighting for a cause that is not beneficial to them. (On a scale from 0 to 10).

A.3.5 Strategic sophistication

We measure strategic sophistication with a simplified beauty contest game (Nagel, 1995). In this game, participants hypothetically play with four other players. Each player reports a number from zero to six. To win the task, the player has to choose a number that is equal to the average of the numbers chosen by the other players minus one. This simple task enables us to identify different types of strategic reasoning:

- If a player thinks that other subjects choose numbers at random, then he or she expects the average number chosen by the other players to be three. The optimal strategy is then choose number two. This corresponds to $k = 1$ behaviour.

[52] The first sure payment is 170 ETB. If the person chooses the sure payment, the next decisions has a sure payment of 80 Birr. If the person chooses the lottery, the next decisions has a sure payment of 260 Birr. These two choices enable us to bound the CRRA coefficient of the respondent. We assign to each respondent the midpoint of the interval of risk aversion consistent with his or her decisions.

- If a player thinks that other subjects are $k = 1$, then he or she expects the average number to be two. The optimal strategy is thus to choose number one. This corresponds to $k = 2$ behaviour.

- Finally, if a player thinks that other subjects are $k = 2$, then they he or she expects the average number to be one. The optimal strategy is thus to choose number zero. This corresponds to $k = 3$ behaviour.

A.4 The Estimation of the Structural Model

A.4.1 The analytical expressions of the moments

A.4.1.1 Noisy selection

Application rates $\Pr\left(C_z \leq \Phi\left(\frac{T_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau\right)$**:**

$$\int\limits_{-\infty}^{\infty} \Phi\left(\frac{\Phi\left(\frac{t_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau-E(C_z|T_z=t_z)}{\sigma_{C_z|T_z}}\right) f(t_z)\,dt_z \tag{A.2}$$

Expected applicant ability $E\left(T_z \mid C_z \leq \Phi\left(\frac{T_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau\right)$**:**

$$\frac{\int_{-\infty}^{\infty} t_z \cdot \Phi\left(Y(t_z)\right) f(t_z)\,dt_z}{\int_{-\infty}^{\infty} \Phi\left(Y(t_z)\right) f(t_z)\,dt_z} \tag{A.3}$$

where

$$Y(t_z) = \frac{\Phi\left(\frac{t_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau-E(C_z|T_z=t_z)}{\sigma_{C_z|T_z}}$$

Dispersion in applicant ability $Var\left(T_z \mid C_z \leq \Phi\left(\frac{T_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau\right)$**:**

$$\frac{\int_{-\infty}^{\infty}\left(t_z - E\left(T_z \mid C_z \leq \Phi\left(\frac{T_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau\right)\right)^2 \cdot \Phi\left(Y(t_z)\right) f(t_z)\,dt_z}{\int_{-\infty}^{\infty} \Phi\left(Y(t_z)\right) f(t_z)\,dt_z} \tag{A.4}$$

where

$$Y(t_z) = \frac{\Phi\left(\frac{t_z-\mu_a}{\sigma_a}\right)(b_z+\tau^w)+\tau-E(C_z|T_z=t_z)}{\sigma_{C_z|T_z}}$$

Expected recruitment probability $E\left(Pr\left(a \leq T_z\right)\right)$**:**

$$\int\limits_{-\infty}^{\infty} \Phi\left(\frac{t_z-\mu_a}{\sigma_a}\right) f(t_z)\,dt_z \tag{A.5}$$

A.4.1.2 Noisy ability

Application threshold c_z^***:**

$$\frac{c_z^*-\tau^w}{b_z+\tau} = 1-\Phi\left(\frac{a-\mu_{T_z}-\frac{\sigma_{T_z}}{\sigma_{C_z}}\rho_z(c_z^*-\mu_{C_z})}{\sqrt{1-\rho_z^2}\sigma_{T_z}}\right) \tag{A.6}$$

Application rates $\Pr\left(C_z < c_z^*\right)$**:**

$$\Phi\left(\frac{c_z^* - \mu_{C_z}}{\sigma_{C_z}}\right) \tag{A.7}$$

Expected applicant ability $E(T_z|C_z < c_z^*)$**:**

$$\mu_{T_z} - \rho_z\,\sigma_{T_z}\,\frac{\phi\left(\frac{c_z^*-\mu_{C_z}}{\sigma_{C_z}}\right)}{\Phi\left(\frac{c_z^*-\mu_{C_z}}{\sigma_{C_z}}\right)} \tag{A.8}$$

Dispersion in applicant ability $Var(T_z|C_z < c_z^*)$**:**

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(t_z - E(T_z|C_z < c_z^*))^2 \cdot \mathbb{I}_E \cdot f(c_z|t_z)f(t_z)\ dc_z\,dt_z}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{I}_E \cdot f(c_z|t_z)f(t_z)\ dc_z\,dt_z} \tag{A.9}$$

Expected recruitment probability $E\left(Pr\left(T_z \geq a \mid C_z = c_z\right)\right)$**:**

$$\int_{-\infty}^{\infty}\left(1 - \Phi\left(\frac{a - E\left(T_z \mid C_z = c_z\right)}{\sigma_{T_z|C_z}}\right)\right) f(c_z)\,dc_z \tag{A.10}$$

A.4.1.3 Parameter estimation and standard errors

We use the formulas reported above to calculate simulated moments for different draws of parameters, and then compute the value of the loss function (9). We minimize this function using MATLAB's unconditional minimizer `fminunc`.

To calculate standard errors, we produce 100 draws of bootstrapped moments (the boostrap procedure includes the estimation of B and the demediation of application rates). We then estimate each of the four versions of the model reported in Table 4 for each set of bootstrapped moments.[53] The standard error of a parameter is given by the standard deviation of that parameter over these replications.

A.4.2 The Internal Rate of Return of the interventions

We calibrate the Internal Rate of Return (IRR) of the interventions to represent the returns to a typical a firm hiring a clerical worker in Addis Ababa. Table A.62 below summarises all our key assumptions, which we describe in detail in what follows. First, we estimate the number of potential applicants. The average firm in our sample receives

[53] To make the bootstrap computationally manageable we truncate the minimisation procedure at 1,500 function evaluations and set a slightly lower optimality threshold. The minimisation procedure is truncated in about 8 percent of the simulations. We only use simulations that have converged to a minimum.

50 job applications for a clerical post. On the basis of this, we assume that the pool of potential applicants is composed of 100 individuals.[54]

Second, we quantify the monthly benefit of the interventions using the following formula:

$$\texttt{monthly benefit} = \pi * \Delta\texttt{E[Raven|hire]} * \texttt{No hires} \tag{A.11}$$

where π is the monetary return to an extra unit of performance on the Raven test, Δ`E[Raven|hire]` is the change in the expected Raven score of a hire, and `No hires` is the number of workers hired. To obtain a value for π, we regress wages on Raven test scores using the data of Abebe et al. (2020). We compute Δ`E[Raven|hire]` using the structural estimates from the noisy-ability case estimated with the core moments.[55] `No hires` is the average number of workers hired by a firm in our sample in a given hiring round. Finally, we assume that the monthly benefit accrues to the firm for 45 months. We obtain this number by taking the average separation rate reported by firms in our survey and calculating the expected duration of a match.

Third, we quantify the one-off cost of application incentives using the following formula:

$$\texttt{cost} = r * \Delta\texttt{E[No Applicants]} + (100 * \texttt{E[No applicants]}) \tag{A.12}$$

where r is the cost of reviewing one additional application. We calculate r using firms' self-reports. In particular, firms report that it takes them about one hour of a manager's time to review an application. We price this hour at the median hourly salary of the HR staff who review applications in these firms. We obtain Δ`E[No Applicants]` (the expected change in the number of applicants compared to the control condition) and `E[No applicants]` (the expected total number of applicants in the incentive condition) using our structural estimates.

Finally, the total cost of the wage intervention is given by (i) a one-off cost of $r *$ Δ`E[No Applicants]` (as the employer needs to review additional applications), and (ii) a salary cost of 1,600 ETB for each hired workers, for three months.

We calculate the Internal Rate of Return (IRR) of the interventions using these calibrated costs and benefits, and applying MATLAB's `irr` function. We bootstrap the

[54]This corresponds to an application rate of 50%, which is in line with the average application rate across treatment conditions on our experiment.

[55]To calculate `E[Raven|hire]`, we first need to calculate the selectivity threshold used by the employer. This is the threshold that ensures that the expected number of applicants with ability above the threshold is equal to the desired number of hires. We then calculate the expected change in Raven scores among applicants above this threshold.

estimation procedure to obtain a confidence interval for the IRR: (i) for each new bootstrap sample, we solve the model again and obtain a new set of parameters, (ii) for each new set of parameters, we calculate the IRR of the interventions, (iii) we use this distribution of IRRs to compute the confidence interval. In step (1), for the incentive and wage intervention, we use the same bootstrapped parameter estimates that we obtained to calculate parameter standard errors. For the counterfactual interventions that target the incentive on the basis of gender or age, we run 100 additional simulations (50 for each model).

Table A.62: Assumptions for the cost-benefit analysis

	Value
Number of potential applicants	100
Number of workers hired	3
Monthly return to one extra point on the Raven test (π)	22.8
Marginal cost of interviewing one more applicant (r)	38.8
Expected tenure on the job (no. months)	45

A.5 Comparison with original plan

For the reduced-form analysis, we follow a registered pre-analysis plan. We have updated the plan in the following way:

1. Heterogeneity. We planned to study heterogeneous treatment effects with respect to individuals' savings, cash on hand, expenditure. These variables were collected during phone call number two. The plan was to ask these questions about the month preceding phone call number one. Due to miscommunication with the field team, this plan was not implemented. Instead, the questions were asked about the last completed month, which in most cases included several weeks after treatment. We are thus unable to use these variables to study heterogeneity. Further, we included three additional dimensions of heterogeneity measured during the second phone call: a measure of credit constraints and two questions on time preferences. The credit constraint question – a newly-designed question applying the logic of multiple price list to the measurement of credit constraints – was hard to understand for respondents according to the reports of the field team. Similarly, the two questions on time preferences were ultimately poorly formulated. These worries compound the fundamental problem that these variables were measured *after* treatment and are thus not suitable to study the heterogeneity of treatment effects. We thus chose not to use them for heterogeneity analysis in the paper.

2. Quantile regressions. Due to a lack of fine-grained variation in the scores at the top and at the bottom of the distribution in the individual tests, we have performed most the the quantile regression analysis on overall indices of ability computed over the pre-specified families of ability measures.

3. Outcomes. We are unable to report results related to the wage paid by and the location of the jobs that jobseekers apply for between the two phone interviews because of the large amount of missing data. For both variables, we have more than 50% missing data.

4. The variable 'value of the job'. This variable was not part of the original plan. This variable has a clear theoretical interpretation and we thus prefer it to splitting the sample based on endogenous stratification, as per our original plan.

5. Experiment two. We originally planned to use experiment number two to obtain the weights that managers placed over the various dimensions of quality. However, we found this challenging to implement. We ran several pilot sessions of

a design where managers were shown several pairs of candidates with different characteristics and had to pick one candidate in each pair. These pilot sessions suggested that managers place a larger weight on cognitive ability, compared to non-cognitive ability or experience. However, managers decisions were very hard to predict and to reconcile with our model of decision utility (our estimated model could predict decisions only modestly better than a random guess). We thus opted for the simpler and more transparent task which is described in the paper.

6. Robustness tests. We have also included several robustness tests that were not pre-specified, but were requested by referees, suggested by seminar audiences or motivated by the findings of the main pre-specified analysis.

A.6 The value of the job

In this section, we describe how we calculate the value of the experiment's job. This value is given by the stream of utility that the worker obtains if they get the job, minus the stream of utility that the worker would have obtained otherwise:

$$b = \begin{cases} V(j) - V(u) & \text{if currently unemployed} \\ V(j) - V(e) & \text{if currently employed} \end{cases} \tag{A.13}$$

where $V(j)$ is the gross value of the experiment's job, $V(u)$ is the value of being unemployed, and $V(e)$ is the value of being employed at the wage that the market currently pays for the worker's skills (we will refer to this as the 'market wage').

We proceed in two steps. First, we characterise $V(j)$, $V(u)$ and $V(e)$ as functions of the wage paid by the experiment's job, market wages, worker impatience and the probability of finding and losing a job. Second, we forecast the market wage of each worker using a Post-LASSO estimator (Belloni et al., 2014) and make informed assumptions about the other parameters. Throughout this section, we assume that time is discrete and measured in months. We also assume that workers have a time-separable, linear utility function of the following form:

$$U_t = \sum_{k=0}^{T} \delta^{t+k} E[w_{t+k}] \tag{A.14}$$

We start by calculating the value of unemployment. We assume that the worker values non-work time at c (Mas and Pallais, 2017). This includes transfers, the value of leisure, etc... We assume that c is given to the worker at the end of the month. Further, we assume that the worker will find a job in the next period with probability p. The value of being in unemployment is thus given by:

$$\begin{aligned} V(u) &= \delta c + \delta^2 p V(e) + \delta^2 (1-p) V(u) \\ &= \frac{\delta c + \delta^2 p V(e)}{1 - \delta^2 (1-p)} \end{aligned} \tag{A.15}$$

The value of being employed, on the other hand, is given by:

$$\begin{aligned} V(e) &= \delta w + \delta^2 (1-q) V(e) + \delta^2 q V(u) \\ &= \frac{\delta w + \delta^2 q V(u)}{1 - \delta^2 (1-q)} \end{aligned} \tag{A.16}$$

where w is the market wage and q is the probability of losing the job in any given period of time. We can substitute $V(e)$ into (A.15) to derive an expression that defines

$V(u)$ only as a function of the parameters c, w, δ, p and q:

$$\begin{aligned} V(u) &= \frac{\delta c}{1-\delta^2(1-p)} + \frac{\delta^2 p}{1-\delta^2(1-p)} \frac{\delta w + \delta^2 q V(u)}{1-\delta^2(1-q)} \\ &= \left(1 - \frac{\delta^4 pq}{(1-\delta^2(1-p))\,(1-\delta^2(1-q))}\right)^{-1} \times \left(c' + \frac{\delta^2 p}{1-\delta^2(1-p)} w'\right) \end{aligned} \tag{A.17}$$

where $c' = \frac{\delta c}{1-\delta^2(1-p)}$ and $w' = \frac{\delta w}{1-\delta^2(1-q)}$. The value of being employed can be obtained by substituting (A.17) into (A.16).

Finally, the gross value of getting the experiment's job for a worker in treatment group f is given by:

$$V(j) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right)$$

The worker will obtain wage w_f for three consecutive months and will then return to unemployment. For simplicity, we assume that work experience in the experiment's job does not affect future wages and that the worker will only hear about new job opportunities in the last month of the job. These assumptions make our estimates of $V(j)$ conservative.

We can now write an expression for the value of the job for an unemployed person. This is given by:

$$V(j) - V(u) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right) - V(u)$$

Further, the value of the job for an employed person is given by:

$$V(j) - V(e) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right) - V(e)$$

In our second step we forecast market wages. To do this, we use the Post-LASSO estimator recommended by Belloni et al. (2014). This estimator is obtained in two stages. First, we regress individual wages on a large set of covariates, using the LASSO estimator and all observations of jobseekers who have a formal job. This allows us to select a sub-set of covariates that can be used for forecasting. Second, we run an OLS regression of wages on the covariates selected in the first stage (using only control group observations, to minimise distortions in reporting potentially induced by the interventions) and use the OLS coefficients to derive a forecast of w for each worker.

The Post-LASSO estimator is recommended to produce forecasts when a large number of potentially informative covariates are available. In these settings, estimators that maximise in-sample fit often have poor out-of-sample properties, as they tend to fit some of the noise in the data. The original LASSO estimator reduces over-fitting by imposing a penalty on non-zero coefficients. More precisely, for a canonical model:

$$y_i = \sum_{j=1}^{p} x_{i,j}\beta_j + u_i \tag{A.18}$$

the LASSO estimator of the parameter vector β is obtained by minimising the following function:

$$\widehat{\beta} = \arg\min_{\beta} \sum^{n} \left(y_i - \sum_{j=1}^{p} x_{i,j}\beta_j \right)^2 - \lambda \sum_{j=1}^{p} |\beta_j|\gamma_j$$

where λ is a penalty parameter and γ_j are penalty loadings. One problem with this estimator is that the non-zero coefficients tend to be biased towards zero. The Post-LASSO estimator reduces this bias by re-estimating the coefficients with OLS.

We use a rich set of variables in order to forecast wages. These variables describe the socio-demographic characteristics of workers, their educational achievements, and their labour market experience. We report the full list of variables in table A.63 below. To maximise the flexibility of our empirical model, we discretise continuous variables and include dummies for each possible discretised value of the variable. Finally, our measure of wages refers to the jobs that subjects held at the time of the first interview.[56] We report the coefficients estimates obtained with the Post-Lasso estimator in Table A.64 below. The first column shows the estimates obtained by using the theoretically optimal penalty and the second column shows the estimate obtained with a manually-set lower penalty, which allows us to capture a number of additional plausible predictors. The predicted values we obtain from these two models are highly correlated. In what follows, we use the predicted values obtained with the optimal penalty.

We make the following assumptions on the remaining parameters. First, we assume that the monthly discount factor is $\delta = 0.786$. To determine this figure, we use the daily discounting factor estimated in a recent experiment in Nairobi (Balakrishnan et al., 2015). The estimates of Balakrishnan et al. (2015) suggest relatively high levels of impatience, which is consistent with the cross-country survey evidence reported by Falk et al. (2016) for sub-Saharan Africa. Second, we set the probability of finding a job to 15.3 percent and the probability of losing a job to 11.6 percent, respectively. These

[56] We windsorise the forecast at the 5^{th} and 95^{th} percentiles so that we do not rely on extreme forecasts. Further, we adjust forecasted wages (by applying a simple location shift) to ensure that the mean of the forecast matches that of representative data for workers in Addis Ababa of comparable age and level of education.

Table A.63: Variables used to forecast wages

Variable	Description
female	Female
age	Age
age_sq	Age squared
born_aa	Individual was born in Addis Ababa
newspaper	Individual has found out about the vacancy in the newspaper
amharic	First language is Amharic
oromo	First language is Oromifa
engineer	Engineering or hard science background
economics	Economics background
social_scientist	Degree in social science (other than economics)
GPA_dummy_	Dummies for GPA score (1 point intervals)
wexperience	Wage work experience (number of months)
wexperience_sq	Wage work experience squared
e_type_	Dummies for type of employer in last job
wage_dummy_	Dummies for wage earned in last job (2,000 ETB intervals)
sexperience	Individual has experience in self-employment
subcity_	Dummies for the subcity of residence of the respondent

figures reflect monthly transition rates from non-employment to employment, and vice-versa, which we calculate using the high-frequency panel data collected by Abebe et al. (2020). Finally, we assume that the value of c is 1,230 ETB. We calculate this figure by using estimates of the value of non-work time from Mas and Pallais (2017) and mean forecasted wages. This figure seems realistic in our context, as unemployed jobseekers report an average monthly expenditure of about 1,000 ETB.

We estimate that the position has positive value for about 61 percent of the individuals in our sample. To confirm that our estimates are informative, we regress the application dummy on our estimate of the value of the job. We find a large and significant correlation: a one standard deviation increase in the value of the job is associated with a 10 percentage points increase in application rates. We report the estimates in Table A.65 below. Finally, in Table A.66 we show that the measure of B that we obtain by using different predicted values of the wage (LASSO forecast with optimal penalty, LASSO forecast with manual penalty, and OLS) are highly correlated with each other.

Table A.64: Post-LASSO regression of wages in Ethiopian Birr (control group observations)

	Optimal penalty	Manual penalty
	(1)	(2)
Heard of job on newspaper	108.186	52.678
	(191.410)	(176.968)
Economics background	719.240	686.784
	(266.857)	(270.642)
Work experience (months)	18.188	16.703
	(3.101)	(3.015)
Worked for private foreign business		179.904
		(1058.206)
Age	45.814	23.520
	(37.495)	(35.126)
GPA dummy (2-3)		239.314
		(165.893)
Previous wage dummy (2000-4000)		995.650
		(519.302)
Previous wage dummy (4000-6000)	1862.386	2091.702
	(438.405)	(437.934)
Previous wage dummy (6000-8000)		4175.537
		(846.208)
Previous wage dummy (8000-10000)	3682.031	4012.018
	(594.769)	(664.741)
Obs.	361	361

Notes: Post-LASSO regressions to forecast market wages. In the first stage of the Post-LASSO procedure, we run a LASSO regression of wages on the set of covariates described in Table A.63. In the second stage, we run an OLS regression of wages on the covariates selected by the LASSO estimator in the first stage. In column (1), we report estimates obtained by applying the optimal LASSO penalty parameter. In the column (2), we report estimates obtained by applying a manually-chosen, lower penalty parameter that enables us to select a larger number of covariates. Robust standard errors in parentheses. Sample used: control individuals employed at baseline.

Table A.65: Regression of applications on the value of the job

	Applied to the experiment's job
	(1)
B (z score)	0.105
	(0.007)
Constant	0.512
	(0.007)
Obs.	4686

Notes: OLS regressions. The dependent variable is a dummy for whether the individual has applied to the experiment's job. The independent variable is the estimate of the value of the job B, obtained using the market wage forecast from the model of column 1 of Table A.64. B is windsorised at the 10^{th} and 90^{th} percentiles. Robust standard errors in parentheses. Sample used: baseline sample.

Table A.66: Correlation between measures of B obtained with different market wage forecasts

	Dep var: B (optimal LASSO)	
	(1)	(2)
B (OLS)	0.919	
	(0.009)	
B (manual LASSO)		0.955
		(0.010)
Constant	-89.205	-74.051
	(8.608)	(4.336)
R2	0.91	0.94
Obs.	3932	4686

Notes: OLS regressions. The dependent variable is the measure of B based on the market wage forecast produced by running the post LASSO estimator with the optimal penalty (reported in in column (1) of Table A.64). This is the measure of B that is used in the rest of the paper. In the first column, we regress this measure of B on an alternative measure of B based on an OLS forecast of the market wage. The OLS estimator uses all the variables that are initially available to the LASSO estimator. In the second column, we regress our main measure of B on an alternative measure of B based on the post-LASSO market wage forecast reported in column (2) of Table A.64. To obtain this second forecast, we impose on the post-LASSO estimator uses a manual penalty parameter. This results in the estimator relying on a larger number of covariates compared to the estimator that uses the optimal penalty parameter. All measures of B are windsorised at the 10^{th} and 90^{th} percentiles. Robust standard errors in parentheses. Sample used: baseline sample.