Good Interviewers Matter*

Tavis Barr[†] Raicho Bojilov[‡] Kate Gautier[§] Lalith Munasinghe[¶]

May 2019

Abstract

Interviews present a puzzle: research suggests they are of little value in screening hires, yet they remain pervasive. Using a longitudinal dataset from a large call center company, we find that interviewers vary in quality along three dimensions of performance: making offers, convincing candidates to join the company, and screening for durable and productive hires. The variation in interviewer performance is larger for the first two dimensions than for screening. Furthermore, interviewers improve over tenure on the first two dimensions, while their ability to screen remains stagnant. Interviewers who are good at making offers are also good at convincing the candidate to accept them, and they do so without sacrificing quality of hire. We find that the company could realize cost savings of over 40 percent of its hiring and recruiting budget by reassigning interviews to their best interviewers. Thus, interviewer performance, especially in often-ignored dimensions, can have a major impact on search and matching of employers and employees.

*We would like to thank P. Cahuc, G. Friebel, D. Huffman, E. Kausel, F. Kramarz, S. Koch, F. Lange, R. Macera, B. MacLeod, and J. Tesada for their helpful comments and suggestions.

†TalenTeck. E-mail: tavis.barr@talenteck.com

‡School of Business and Management (Escuela de Administración), Pontificia Universidad Católica de Chile; CREST-Ecole Polytechnique. E-mail: raicho.bojilov@uc.cl.

§TalenTeck. E-mail: kate.gautier@talenteck.com

¶Department of Economics, Barnard College and School of International and Public Affairs; Columbia University. E-mail: lm25@columbia.edu.

1 Introduction

A long and evolving literature questions the effectiveness of interviews in screening for quality hires. Yet the continued pervasiveness of interviews must imply that firms find value in conducting them.

We shed light on the source of this value using a large dataset that includes detailed information on interview, hiring, and employment outcomes in a call center company. Our results suggest that researchers may have overlooked an important function of interviewers by focusing simply on the single dimension of screening: Good interviewers may only be a mildly effective means of screening productive workers, but they play a crucial role in convincing candidates to accept job offers. By getting a higher percentage of approved candidates to accept their job offers, effective interviewers are able to generate a larger pool of equally productive workers from the same pool of candidates. Thus, the detection and deployment of high-quality interviewers can be a critical lever for improving outcomes in the job search process.

Our dataset tracks a large number of interviewers over several years of conducting interviews, in a setting where candidates are assigned to interviewers randomly. We observe outcomes at each stage of the hiring process – the firm's offer, the candidate's acceptance, and their post-hire performance. The firm's offer depends almost entirely on the recommendation of a single interviewer, allowing us to link outcomes directly to an interviewer's decisions. These unique data enable us to define interviewer "skill" across three dimensions: the ability to fill seats (job offer rate), to sell the job to prospective candidates (acceptance rate of a job offer), and to screen for durable and productive hires (post-hire stay rate). Because these outcomes appear successively, we jointly estimate the offer, accept and stay decisions using a sequential choice model. Finally, a rich set of interviewer and interviewee characteristics allows us to establish patterns in both observable and unobservable heterogeneity across interviewers.

Our results show that there is substantial, time-persistent heterogeneity in performance between interviewers. We find that the heterogeneity of interviewers' ability to fill seats and to sell the job is substantially larger than that of their ability to screen. Despite observing a rich set of interviewer characteristics, we find that little of the heterogeneity of interviewer performance is explained by them. This implies that selecting for interviewer quality on the basis of observable dimensions may be difficult; however, the persistence in skill over time suggests firms may be able to use early performance as a means for identifying their best interviewers.

We then determine how these three skill dimensions are related to each other. Our results show that the interviewer's individual propensity to extend offers is positively correlated with the ability to sell the job, and that neither of the two is significantly correlated with the ability to screen. In other words, it is not the case that interviewers who are able to fill more seats do so by lowering their standards. Rather, good interviewers are able to generate more hires of the same quality than poor interviewers who face the same candidate pool.

Next, we investigate how interviewers' performance on these dimensions changes over time. We find that interviewer abilities to fill seats and to sell the job improve significantly with job tenure, while the ability to screen is relatively stagnant. Interestingly, the rate of improvement depends on the interviewer's work history at the firm. The tenure effect on the offer and accept decisions is largest for interviewers in the Human Resources division who have previously worked as regular call center agents.

Given these skill differentials across interviewers and the associated learning dynamics, the question of allocation becomes important. We find that the firm is actually offering more job experience to the interviewers who are worse at filling seats and screening. The result is that the worse interviewers end up catching up somewhat with the better interviewers through on-the-job learning. Our modeling approach allows us to investigate the implications of a better allocation of interviewers. We find that by re-assigning interviews to its best interviewers – thereby giving those interviewers both a chance to work their magic and a chance to learn more on the job – the firm could realize cost savings of over 40 percent of its hiring and recruiting budget, which amounts to nearly 2 percent of operational profits. Hence, we conclude that the heterogeneity in interviewer skills is not just statistically significant but also meaningful in a business context.

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 presents the data. Section 4 conducts descriptive and non-parametric analysis to establish the presence of heterogeneity in hiring and employment outcomes across interviewers. Section 5 presents our empirical framework, while section 6 discussed our main empirical results. Section 7 considers some robustness checks related to performance and interviewer attrition. Section 8 quantifies the impact of interviewer heterogeneity on hiring, turnover, and profits, and Section 9 concludes.

2 Literature Review

Given their long-standing centrality to the job search process, employment interviews have attracted widespread research for a long time (early examples include McMurry (1947) and Glaser, Schwartz and Flanagan (1958); see also McDaniel, Whetzel, Schmidt and Maurer (1994) for a broad review to that date). Some general patterns have emerged from that research, though by no means unanimously. McDaniel et al. (1994) and Schmidt and Zimmerman (2004) show that structured interviews are more effective than unstructured interviews. According to Maurer and Lee (2000), having more interviewers may increase the accuracy of the interview as a screening mechanism, though Fific and Gigerenzer (2014) find otherwise. Dipboye, Gaugler, Hayes and Parker (2001) argue that some interviewers are more effective than others. A key finding in recent research (see: Dana, Dawes and Peterson (2013); Kausel, Culbertson and Madrid (2016), and Hoffman, Kahn and Li (2017)) is that interviewers greatly overestimate their effectiveness at screening candidates. In fact, Hoffman et al. (2017)) and especially Cowgill (2018) show that machine learning algorithms can improve upon the screening outcomes achieved by interviewers.

We find suggestions in the business literature that employment interviews may have value beyond their ability to screen candidates, such as selling the firm to the prospective employee and making team members feel responsible for the success of any resulting hire; see for example Knight (2015). Indeed, given their continued prevalence in the hiring process, we are inclined to think that interviews provide value to the firm. Nevertheless, as far as we are aware, academic research has had a singular focus on evaluating interviews as a screening mechanism. We are not aware of any studies to date examining the effectiveness of interviews at achieving other firm goals, and we believe that our research is a first in this vein.

A second contribution of our work is that while research into interview effectiveness has exhibited remarkably detailed measurement and even experimental variation of the content of the interview, relatively less has been studied about the impact of background and work experience on the effectiveness of interviewers as workers. Among the known findings are that training increases interviewer effectiveness (see Huffcutt Woehr (1999)) and that older interviewers intimidate younger interviewees (see Allen, Schetzsle, Mallin, Pullins (2014)). Unlike most of this work, we not only have some information in our data about the content of the interviews, but also very rich longitudinal data about the employment history of the interviewers and the hiring and employment outcomes of job applicants.

Finally we note that recent research raises the riveting possibility that the hiring functions of human resources departments may be replaced by machine learning algorithms. Hoffman et al. (2017) compare decisions of hiring managers to those of an algorithm that is based on a test and other application data, and find that the algorithm outperforms hiring managers. Cowgill (2018) has similar results but focuses on pre-interview screening rather than the final hiring decision. Although in our context candidates do undergo an algorithmic-based pre-screening prior to being interviewed, our primary focus is not on comparing human screeners against such an algorithm, but rather against each other. To this end, we examine how both observable and unobservable interviewer characteristics influence hiring outcomes, characterize the heterogeneity in interviewer performance, and quantify its value.

3 Data

Our data come from the US-based call centers of a large multinational company whose main activity is the collection of debt. We focus on operations in which the outstanding balance is small relative to the median income in the US and does not vary much across accounts. We have applicant records for every candidate in our study period, which track the full progress of every applicant from the initial online application through the interview process to the actual start of work, as well as personnel records for each employee, which allow us to track hiring and employment outcomes for any individual who applied for a job at the company. Formal interviews are an integral part of the hiring process, and the associated records can be linked with other individual-level information available in the applicant and personnel records of the company. Finally, we also have access to a systemized record of the comments provided by the interviewers after the completion of the interview.

3.1 The Hiring Process and Job Interviews

The company instituted the current version of its interview program in the beginning of August 2012. As part of the hiring process, job applicants fill in a questionnaire during an initial online application. An algorithmic technology similar to the one in Hoffman et al. (2017) uses the associated responses to generate a score, which in combination with local labor supply and demand shocks determines a score threshold above which candidates get selected for an interview. The firm conducts one-on-one phone interviews and it retains a full record of all contact attempts, the identity of the interviewer, and the hiring recommendation of the interviewer, along with an assessment by the interviewer for each completed interview.

The interviews are conducted in a structured manner and interviewers are requested to collect information during the course of the conversation about: the speaking skills of the applicant, the listening skills of the applicant, the preference of the applicant for flexible hours, and an evaluation of the professional level of the applicant, expressed in terms of a prospective initial job position within the hierarchy of the company. Importantly, the evaluation by the interviewer remains confidential information that is not shared with the job applicant. Upon the completion of the interview, the interviewers fill in an electronic form that systematizes their interview impressions and records their hiring recommendation. In the overwhelming majority of the cases, the firm makes an offer decision after one interview. In about 8 percent of the cases, however, there are one or more additional interviews with the job applicant. In all cases, those who conduct the interviews also make the offer decisions. If there are multiple interviews, the offer decision is made after a discussion between all interviewers. The information from the additional interviews is aggregated to reach the final offer decision of the committee. In our work, we limit ourselves only to the first interview of each candidate. The results do not change if we drop out the applicants who have more than one interview.

HR managers and HR agents conduct most interviews, followed by the team managers and then senior call agents who get drafted intermittently to help out with the interviewing process. There are neither official nor unofficial rules about the matching of applicants to interviewers: Who does a specific interview is primarily a question of availability. It seems that the only rule of thumb is that the management does not want to overburden individual team managers or senior level call operators with the additional task of doing interviews, so it tries to spread the assignment as evenly as possible across capable employees.

The approved applicants receive job offers, mostly within two weeks of the interview. We consider an offer to be accepted if an applicant starts working at the company. Those who accept the job offer attend a two week training program. As part of the training program, the newly hired individuals get introduced to the work, take some obligatory courses related to their and the company's legal obligations and rights, and pass an exam to certify that they have a good understanding of these matters. Importantly, all newly hired workers go through the same job training after which another set of team managers decides whether some individuals merit a fast track promotion.

Most new employees enter the company at the bottom of the hierarchy. In that sense, fast track promotions are rare. Therefore, the interviewer's evaluation of the professional level of the job applicants does not drive promotions. More than 95% of all employees work full time. Each workstation consists of a computer, a telephone, and a recording device. Importantly, only one worker handles a given call and an automatic switchboard assigns inbound and outbound calls by matching the call at the top of the queue with the longest waiting operator, making the distribution of tasks effectively random across workers.

Compensation consists largely of a base pay linked to the hierarchy level and it is comparable to that in the manufacturing sector: workers of tenure longer than six months receive between 14 and 21 dollars per hour. Promotions are closely related to worker performance, which is measured on a scale from 0 to 5. Those who receive 3 or more are considered for promotion to the next hierarchical level. Importantly, the performance measures are completely unrelated to the recruitment process or the people involved in it, while promotion decisions are made by a committee of supervisors who are generally not involved in the interview and hiring decisions of potential subordinates.

3.2 Data Description and Summary Statistics

The data that we use in our analysis covers the period from July 31, 2012 to August 1, 2017. We limit our analysis to US call centers of the company where interviews have been conducted by at least 10 different interviewers. To limit small sample size concerns when we estimate the effect of individual interviewers on applicants, we consider only job applicants who have been interviewed by people who conduct at least 200 interviews in total. We have further limited our analysis to the subsample of those applicants interviewed by employees who conducted their first interview after July 31, 2012 in order to have a sample of interviewers with complete work history. This leaves us with a sample including data on 88 interviewers and 129,636 job applicants, of which 56,651 have been interviewed.

Table 1 shows that 43 percent of applicants receive an interview, while 54 percent of interviewees receive an offer, implying that only 24 percent of all applicants get an offer. Furthermore, the probability that a job offer is accepted is around 65 percent. Thus, the company hires only about 15.5 percent of all applicants. During the first six months on the job, the monthly hazard rate of separation declines and eventually levels off. Out of all hires, the share of workers who remain employed by the company for at least six months is around 31 percent. In other words, only a bit more than 5 percent of all job applicants are hired and retained by the company for more than six months. Because of this high early separation rate, and the heavily front-loaded cost of the employee's initial two-week training, turnover is a major concern for the company, as it is the largest source of operational inefficiency. For this reason, we consider the separation decision during the first six months of employment as our primary outcome variable of interest.

Table 2 describes the characteristics of the interviewers at the company. The average number of interviews is 2022, but the median is only around 560. The large variance in the number of interviews highlights a great heterogeneity in the total number of interviews across interviewers. The great dispersion in the number of interviews per interviewer remains even after we control for location differences. The average tenure as an interviewer in the sample is 15 months, and on average people conduct 5 or 6 interviews per day. We also find that there is considerable variation in the number of interviews per day, with the absolute maximum of interviews per day at 26. Around 80 percent of the interviews are conducted by women, and about 37 percent by people from traditionally disadvantaged ethnic or racial backgrounds. About 13 percent are conducted by an HR manager or team manager, while about 37 percent are conducted by firm employees who belong to the HR division. The data in the table also show that many call operators eventually join the HR division. Movement in the other direction is not observed.

Before we proceed with the analysis of the data, we quickly verify that the firm allocates job applicants to interviewers in an essentially random manner. Table 3 presents the correlations between the characteristics of the interviewers and the job applicants.

To control for differences in local labor market conditions, we subtract from the individual characteristics the corresponding average by location. None of the resulting correlation coefficients is significant or large in magnitude. Thus, we conclude that the assignment of interviewers to job applicants is essentially random. This finding conforms to the anecdotal evidence that we have obtained during our interaction with the representatives of the company.

4 Descriptive and Non-Parametric Analysis

In this section, we establish in a descriptive manner that there is heterogeneity in applicant outcomes across interviewers. We apply an approach to the study of persistent differences that can be traced back to Pakes and Ericson (1999). Specifically, we condition on the history of hiring and employment outcomes of applicants across interviewers and show that this history is predictive of subsequent hiring and employment outcomes by the same interviewer.

Figure 1 illustrates how the distribution of average job offer, accept, and stay rates of job candidates differ across interviewers. The three charts in the top row show that, although interviewees are essentially drawn randomly from the same pool of candidates, the raw rates vary across the whole range from 0 to 1. However, there are two key confounding factors creating heterogeneity that is not due to interviewer differences. The first is location: As mentioned above, assignments are quite random within location, but applicant characteristics vary across locations. For this reason, charts 1.1-1.3 plot also the distributions of interviewer offer, accept, and stay rates, controlling for local labor market conditions by subtracting the location-specific variation in the corresponding rates. They show that the previously documented heterogeneity in the interviewer rates is not driven solely by differences in local labor market conditions across different call center locations.

The second confounding factor is related to differences in experience and professional background. For this reason, charts 2.1-2.3 compare the offer, accept, and stay rates for HR and non-HR employees. The main difference between these two groups is that HR employees tend to make many more offers than non-HR employees. However, this high offer rate does not translate into a high accept or stay rates. Overall, the differences between HR and non-HR interviewers, such as they are, cannot account for the observed heterogeneity in hiring and employment outcomes of job candidates across interviewers. Next, charts 3.1-3.3 present the distributions of interviewer offer, accept, and stay rates for interviewers with more than 1000 interviews. The distributions of the offer and stay rates are similar to those discussed above. The only substantial difference is in the offer rates. Unlike the distribution of the offer rates for all interviewers, the interviewers who have conducted more than 1000 interviews have on average a higher offer rate. Nevertheless, substantial heterogeneity in outcomes remains across interviewers.

We continue by calculating the average offer, acceptance, and stay rates for the first 50 interviews by interviewer. To control for local labor market effects, we normalize the individual interviewer rates by subtracting the corresponding location averages. For this reason, the normalized interviewer rates are centered at zero and bounded between -1 and 1. Then, we investigate whether the offer, acceptance and stay rates for the first 50 interviews predict the respective offer, acceptance and stay rates for the following 50 interviews. The results from this exercise are reported in Figure 2. On the horizontal axis we plot the average offer, acceptance, and stay rates for the first 50 interviews. The vertical axis of the charts in the first column represents the corresponding offer, acceptance and stay rates for the following 50 interviews. The second column of charts investigates the relationship between the rates for the first 50 interviews and the corresponding rates for interviews 150-200, while the third column investigates the relationship between the rates for the first 50 interviews and the respective rates for interviews 450-500.

As we can see, there is a strong positive correlation between the offer rate for the first 50 interviews and the offer rates for subsequent interviews. Moreover, we find that the linear prediction of the relationship between the past and future offer rates is always above the 45-degree line: This finding suggests that interviewers "learn" to make more offers during their first 150-200 interviews. Furthermore, the predictive power of the first 50 interviews erodes little as we try to predict more and more distant offer rates in the future. Thus, the data provide evidence for persistent differences in offer rates across interviewers.

Next, we consider whether there are similar persistent differences in the acceptance rates of job offers across interviewers. We interpret the effect of interviewers on job acceptances as ability to "sell" the company to a job applicant whom the firm wants to hire. The relationship between past and future acceptance rates is very similar to that for offer rates. We find that the acceptance rate for the first 50 interviews positively predicts the acceptance rate for the following 50 interviews, as well as subsequent batches of 50 interviews. Again, the relationship seems to be stable over time, suggesting the presence of time-invariant heterogeneity across interviewers on this dimension. We find that these predictions also lie above the 45-degree line, meaning that interviewers "learn" how to convince desirable job applicants to accept the company's job offer.

Finally, we turn our attention to differences in stay rates across interviewers. The results show again that the interviewer stay rate based on the first 50 interviewed applicants positively predicts the stay rates of subsequently interviewed applicants. Examining Charts 3.1-3.3, we find that the positive relationship remains stable as we consider more and more distant future stay rates. Unlike the cases of the offer rate and the acceptance rate, the linear prediction in the charts is clustered around the 45-degree line, not above it. Thus there appears to be on average no learning by interviewers about how to screen applicants with lower turnover potential.

5 Estimation

We have just established that interviewers vary in their ability to screen and attract candidates who will accept an offer and stay with the company. We now wish to explore the drivers of this heterogeneity, which may include fixed interviewer characteristics (e.g., gender), time-varying interviewer characteristics (most especially job-specific experience), and company decisions about which interviewers gain experience etc. To do so, in this section, we develop a selection choice model of the offer, acceptance, and stay decisions of the company and applicant, conditional on an applicant being selected for an interview. The variables in our model consist of information available at the time individuals apply for a job at the company, along with the expected dates of various hiring milestones as of the moment of application, and a multi-dimensional interviewer rating of the interviewee. Any other information obtained by the firm or applicant in the course of the hiring process, in particular during the interview, remains unobservable to the econometrician. The differences and the overlaps in the information sets of the interviewer, the applicant, and the econometrician may induce correlations between the error terms in the equations that we have used to represent hiring decisions and employment outcomes. Without controlling for the resulting selectivity on unobservables, single-equation methods lead to biased estimates. For these reasons, we jointly estimate the equations associated with the decision to make an offer, accept a job offer, and separate within six months of the beginning of the employment relationship.

Let subscript i denote observations associated with applicant i. We define X_i^k to be a vector of characteristics observable to the econometrician that impact outcome k, where $k = o, a, s$ stand for offer, acceptance, and stay. Let θ_j^k be the effect that interviewer j has on job applicants and 1_{ij} is a dummy that is equal to 1 if applicant i is interviewed by j and is 0 otherwise. The following sequential choice model is taken to the data. First, the firm decides whether to make an offer to applicant i:

$$o_i = 1\left[X_i^o\beta^o + \sum_{j=1}^{J}\theta_j^o 1_{ij} + \varepsilon_i^o > 0\right] \tag{1}$$

If the applicant receives an offer, she chooses whether to accept it:

$$a_i = 1\left[X_i^a\beta^a + \sum_{j=1}^{J}\theta_j^a 1_{ij} + \varepsilon_i^a > 0\right] \tag{2}$$

If the offer is accepted, the worker decides whether to stay for more than six months, the duration of the transitional dynamics before the monthly separation rate levels off:

$$s_i = 1\left[X_i^s\beta^s + \sum_{j=1}^{J}\theta_j^s 1_{ij} + \varepsilon_i^s > 0\right] \tag{3}$$

Formal discussion of identification conditions can be found in Maddala (1983) and Heckman and Navarro (2007).[1] The selection on unobservables is identified from the statistical dependence between hiring choices and employment outcomes, controlling for observable characteristics. For example, the correlation between the errors in the offer and stay equations can be identified from offers and subsequent stay decisions, conditional on observable characteristics.[2] As usual in choice models, it is impossible to identify scale and location parameters, so the standard normalization applies: $\varepsilon_{ir}^k \sim N(0,1)$ for $k = o$, a, s. To achieve nonparametric identification, there must be at least

[1] Heckman and Navarro (2007) focus on the identification of information sets and treatment effects in the presence of dynamic selection.

[2] See, for example, Heckman and Navarro (2007) for more details.

one variable that affects offer but not subsequent outcomes, and another variable that affects acceptance but not stay and performance. In other words, X_i^s is a strict subset of X_i^a, which is a strict subset of X_i^o. We postpone the discussion of our exclusion restrictions until the presentation of the empirical results. To estimate this multi-stage model, we use simulated maximum likelihood (SML) based on the Geweke-Hajivassiliou-Keane smooth recursive conditioning simulator. To save computing time, we generate Halton draws for the SML.

6 Results

We include as explanatory variables the characteristics of the applicant, time-varying characteristics of the interviewer (e.g., interviewer tenure), interactions between the characteristics of interviewers and interviewees, and dummies to capture the time-invariant interviewer effect on outcomes, as well as time and location dummies. The results from estimating the baseline model are reported in Table 4. We first discuss the time-invariant and time-varying interviewer effects and then switch to the impacts of formal evaluations by the interviewer and of the interactions between interviewer and interviewee characteristics. We conclude by reviewing the evidence for selection on unobservables.

6.1 Interviewer Time-Invariant Effects

6.1.1 Estimates

Figure 3 explores the impact of time-invariant interviewer effects on job offers, job acceptances, and stay decisions. Charts 1.1-1.3 show the distribution of these characteristics when they are sorted in increasing order (essentially an inverted cumulative distribution function). In all three equations the dummy associated with an arbitrarily chosen interviewer "0" is omitted. In other words, the impact of interviewer "0" is normalized to 0 in all three equations. The graphs capture the difference in the effect of each interviewer relative to the impact of interviewer "0". The charts show the point estimates for each interviewer, along with the associated 95% confidence intervals. As can be seen, there is a large variation in the interviewer effects on offer decisions. These effects are precisely estimated and there are clear and large differences in the impact of different interviewers on offer decisions. In particular, the overwhelming majority of the interviewer offer effects are distinct from the impact of the reference interviewer. Relative to the interviewer offer effects, the estimated interviewer accept effects exhibit slightly lower variation. But, like the effects on offers, the effects on job acceptance are precisely estimated and there are clear differences in the ability of different interviewers to affect the acceptance decision of job applicants. Finally, we find also that there are statistically significant differences in the impact of different interviewers on stay decisions. However, the variation is much smaller.

Table 5 presents the formal likelihood ratio test for the joint significance of the interviewer offer, accept, and stay effects. It shows that the hypothesis that all interviewer effects are zero in all equations is rejected at the 0.1% significance level. Similarly, we reject the hypotheses that only the interviewer offer effects are all zero, that only the interviewer accept effects are all zeros, and that only the interviewer stay effects are all zeros. The last test result is somewhat surprising, given what we know from Chart 1.3. Nevertheless, the related likelihood ratio test even rejects the hypothesis that all but the top and bottom 10 percent of the interviewer stay effects are equal to zero.

Charts 2.1-2.3 of Figure 3 plot how the probability of offer, accept, and stay for the average candidate vary across the time-invariant interviewer effects. If interviewer effects were all zero, then there would have been no differences in the likelihood of the outcomes across interviewers. The charts show that offer and accept decisions vary widely across interviewers: the range of the corresponding distributions covers almost the whole interval between zero and one. In contrast, the range of the probability of stay across interviewers covers only about a quarter of the interval between zero and one. Thus, we conclude that fixed interviewer characteristics produce a lot of heterogeneity in the hiring outcomes for the average candidate across interviewers but much smaller heterogeneity in employment outcomes across interviewers.

The next issue that we explore is the relationship between the interviewer effects in the offer, accept, and stay equations. Charts 3.1-3.3 provide scatter plot illustrations of the correlation effects reported in the bottom panel of Table 4. Namely, we find a positive but only borderline statistically significant correlation between the interviewer offer and stay effects. Similarly, we find no significant correlation between the acceptance and stay equations. The only correlation that is positive and highly statistically significant is the correlation between the interviewer offer and accept effects. Thus, interviewers who are able to convince candidates to accept a job offer, or who work hard to achieve that, also tend to make more job offers.

6.1.2 Explaining Interviewer Effects

Having established that there is time-invariant interviewer heterogeneity along some dimensions, we might wonder whether interviewer quality can be easily predicted. Table 9 attempts to shed some light on the sources of the estimated interviewer effects and their relationship to observable interviewer characteristics. In the first regression reported in the table, we try to evaluate how much of the interviewer offer effect can be explained by individual interviewer characteristics. The results show that women make fewer job offers on average than men, while HR and team managers tend to make more job offers than other types of interviewers. No other time-invariant characteristics have a significant coefficient in the regression. Next, we investigate the relationship between the interviewer accept effects and interviewer time-invariant characteristics. The regression results show that black interviewers tend to generate more acceptances, along with interviewers who work in the HR division of the company. In contrast, no individual interviewer characteristic can predict the interviewer stay effect.

In the last column, we regress the residual from the regression of interviewer offer effects on stay and accept effects. This produces the part of the interviewer offer effect that is completely unrelated to the impact of the interviewer on job acceptance and stay. Thus, the residual can be interpreted as a proxy for the idiosyncratic bias of the interviewer when deciding how generous to be with job offers. Interestingly, we find that women are more likely to have a negative offer bias than men, while senior level interviewers are more likely to have a positive bias in their offer decisions.

Our main conclusion is that it is hard to predict who is a good interviewer on the basis of easy-to-observe characteristics. Even in the case of offer and accept effects, we find that observable interviewer characteristics account for only 20-30 percent of the variation of interviewer effects. Furthermore, we find that much of the variation in the propensity of interviewers to make offers is driven by idiosyncratic considerations that are independent of subsequent candidate outcomes or observable characteristics. In this sense, our findings confirm results in the preceding literature that a large part of what drives interviewer offer decisions is wholly unrelated to the quality of the prospective employer-employee match.

6.2 Time-varying Interviewer Effects

The interviewer's employment history and tenure as interviewer comprise the time-varying interviewer characteristics available for us to examine. The results in the top panel of Table 4 show that tenure has a significant impact on offer, acceptance, and stay decisions. Furthermore, they show that the impact of tenure varies significantly with the work history of the interviewer within the company. In particular, we investigate the impact of working at the HR division, starting work in the company directly in the HR division, the impact of being a team manager or HR manager, and the impact of starting work at the company straight at one of these senior positions. We include a common tenure term, and additional interactions between tenure and indicators for each of these four categories of interviewers. The omitted category is interviewers who started and continue working as call agents. The common tenure term does not affect offers, but it has a strong positive and statistically significant impact on acceptance and stay decisions. At the same time, the estimates reveal that interviewers who started as HR or team managers tend to make more offers as they gain tenure. Interestingly, individuals at senior positions do not improve over time in their ability to generate acceptances. The most interesting finding with respect to tenure is that interviewers who started as regular workers but eventually joint the HR division tend to improve over time in their ability to screen for applicants who are likely to stay for a long time in the company. In contrast, interviewers who always worked in the HR division do not improve their screening skills with tenure. This result suggests that it is beneficial if interviewers have personal work experience at the job position for which they are conducting interviews. Table 4 also shows that the number of past interviews on the same day has a statistically significant negative impact on offer and job acceptance, but the impact is of relatively small economic importance.

The number of similar tasks performed in the past is a classical measure of learning-by-doing or accumulation of experience, more broadly. The results in Table 4 show that the impact of the number of past interviews on current performance as an interviewer is somewhat limited: Having conducted more interviews leads only to higher offer rates.

Figure 4 investigates further the impact of time-varying interviewer effects (interviewer experience from now on) on offer, acceptance, and stay decisions. Charts 1.1-1.3 plot the point estimates of the impact of interviewer experience on the last recorded offer, accept and stay decision of that interviewer in the database. Thus, the estimated effects represent the maximal impact of interviewer experience. The range of the interviewer experience effects on offers is only marginally less than the range of the time-invariant interviewer effects reported in Figures 3. The results also show that in more than half of the cases, interviewer experience has a positive impact on offers at the 5% significance level. Similarly, we find that interviewer experience has a positive impact on accept decisions and that the maximal impact on accept decisions is similar to that of the interviewer time-invariant accept effect. Again, more than half of the estimated effects are significantly different from zero at the 5% significance level. In contrast, Chart 1.3 shows that interviewer experience has limited or no effect on stay outcomes. Overall, we conclude that the largest impact of the time-varying interviewer effects is at most equal to the impact of the time-invariant interviewer effects.

Charts 2.1-2.3 plot how the probability of offer, accept, and stay for the average candidate vary with interviewer experience. In more detail, we compute the "marginal effect" ofi interviewer experience on hiring and employment outcomes of the average candidate when interviewed by our reference interviewer "0." The charts show that offer and accept probabilities vary widely with interviewer experience, especially in the case of job acceptances. In contrast, the range of the probability of stay across interviewers is quite restricted. Thus, we conclude that there is a lot of heterogeneity in the hiring outcomes for the average candidate across interviewer experience. In contrast, the probability of staying varies little with interviewer experience.

Table 4 shows that the effect of interviewer experience on hiring and employment outcomes depends mostly on the interviewer's tenure in the company. Charts 3.1-3.3

investigate this issue further by plotting the impact of interviewer tenure on the offer, accept, and stay probabilities for the average job candidate. Chart 3.1 shows that tenure has the highest positive impact on the probability of making an offer for interviewers in the HR division who have never worked outside of the HR division. This effect is especially strong for HR managers. In contrast, we find that tenure has the largest positive effect on accept and stay decisions for interviewers who work at the HR division but started elsewhere in the company.

6.3 Interviewer Comments

One of the great advantages of our data is that we have both a systematized record of the impressions of the interviewers and detailed information on hiring and employment outcomes. In our analysis, we consider how the impressions of the interviewers about the speaking skills, listening skills, the applicant's preference for flexible hours, and the relative professional aptitude of the candidates affect job offers, job acceptance, and stay decisions. The associated variables were included in the baseline specification but most were not reported in Table 4 beacuse they were insignificant. The most important result is that the subjective evaluation by the interviewer of the professional aptitude of the candidate has a very strong predictive effect on hiring and employment outcomes. Specifically, if the interviewer concludes than an individual would be able to perform the tasks associated with more senior job positions, then the candidate turns out to be both more likely both to accept an offer and to stay longer.

As discussed in the description of the data, the interviewer's comments on the topic are not shared with the job applicant and the people responsible for promotions are different from the interviewers. Importantly, we obtain this result while controlling for the observable characteristics of the applicant, which implies that the interviewer's comments incorporate (private) information that is hard to observe. Thus, we interpret this finding as evidence that the interviewer's inference about the aptitude of the job applicants helps improve both screening and self-selection during the hiring process.

6.4 Interactions of Characteristics of Interviewers and Applicants

We also consider whether gender, ethnic and age differences between the interviewer and the applicant affect hiring and employment outcomes. The top panel of Table 4 shows that gender and ethnicity differences play no significant role at any stage of the hiring process. However, the estimates reveal that interviewers tend to penalize candidates who are more than 5 years younger than them, while they seem to be positively predisposed to candidates who are more than 5 years older than them. Age differences have similar effect on job acceptances. Nevertheless, we do not find that age difference has any impact on the applicants' probability of staying in the company for more than six months.

6.5 Selection on Unobservables

The correlations between the errors in the offer, acceptance and stay equations are reported in the first row of the bottom panel of Table 4. The first row of this bottom section shows the errors at the interview level. However, the other correlation coefficients are not statistically significant. We interpret these results as evidence that on average interviewers are capable of convincing or screening along hard to observe dimensions for candidates that are likely to accept a job offer and that they make offers to these applicants. Naturally, alternative explanations are also possible. The key takeaway is that there is selection during the interview process on hard-to-observe but outcome relevant dimensions. Furthermore, if ignored, this selection on unobservables would lead to biased estimates.

6.6 Exclusion Restrictions

Non-parametric identification can be achieved through the exclusion restrictions discussed below. The first two exclusion restrictions are based on the time between the date of completing the test, on the one hand, and the date on which the interview takes place and the date on which the offer decision is generated, on the other. This waiting period depends on the day of the hiring cycle on which a given applicant takes the test. Since the average waiting time is around two weeks, the time to offer likely has an impact on extending and accepting a job offer through the likelihood that an applicant still has not accepted an alternative job offer during the current job search. However, it does not affect the prospects of remaining employed in the firm in the long run. Crucially, the schedule of the hiring cycle is not public. Table 4 shows that both of these variables have a significant negative effect on both job offers and job acceptances.

The second exclusion restriction relates to the technological process at the firm and the way the firm determines who is to be interviewed. Soon after the diagnostic entry test, the employer decides who receives an invitation to an interview. This decision is based on a comparison between the applicant's test score and a historically determined threshold. The technical explanation is that the distribution of scores and the associated ranking for each wave of applicants is aggregated and becomes known to the recruiters only by the time the firm makes offer decisions or, sometimes, even post-hire. Thus, we include in our model two variables that affect interview and offer decisions but not subsequent outcomes: the average test score of individuals interviewed in the past 90 days before the candidate applied.[3] Undoubtedly one's test score and ranking relative to the other current applicants affect decisions during the hiring process and, thus, we include them as controls.[4]

Based on the observed evidence, the day on which the applicants take the test affects the probability of getting an offer. One possible explanation why that may be the case is related to organizational pressures to achieve weekly targets related to the number of interviewees and offers. In any case, the estimated effect of the impact of the day of the week on offers is significant. In particular, we find that candidates who fill in the application during the weekend are less likely to get an interview and eventually a job offer.[5]

7 Robustness Checks

7.1 Performance

We count a hire as successful when the employee is retained for more than six months, because one of the main objectives of this firm's management has been to reduce separations. Still, we would like to investigate whether interviewers have an impact on employee performance. Unfortunately, the relatively low hiring rate and high turnover imply a very small sample of employees with performance evaluations: there are only 12 interviewers who interviewed more than 50 future employees who ended up staying

[3]In practical terms, we explore several time horizons: interviewed candidates in the preceding 30, 60, and 90 days before a specific individual applies for a job.

[4]In our preliminary work, we did not find that the characteristics of the other employees in the company have any effect on the probability that a particular individual receives an offer, accepts it, stays long, and performs well. These findings conform to what we know about the operations of the company and the industry. Similarly to the rest of the industry, the biggest concern of the company is filling in its positions and reducing turnover.

[5]In a preliminary run, we tested whether the day of the week when the test was completed has an impact on job acceptance: the results showed no significant impact.

for at least the requisite six months in the company. This leads to two types of sample selectivity concerns: First, for this restricted subsample, we have concerns about the identification of the impact of interviewers from the impact of local labor market conditions. Second, the estimation of a selection model with a performance equation as a final stage on the larger set of interviewers may be biased due to incidental parameter problems.

To circumvent these data limitations, we estimate a modified version of the model in which our measure of a successful hire is that the employee both stays for more than six months and has performance higher than the company's minimal standard. All other hires are considered failures. The comparison of the results from this modified specification and the benchmark specification allows us to investigate whether attracting employees who are likely to stay in the firm comes at the expense of reducing performance quality.

The results from the estimation of the modified selection model are reported in Table 6. There are only two substantive differences relative to the estimates of the baseline model. Under the modified specification, the number of prior interviews conducted by someone working at the HR division, who started elsewhere in the company, has a positive impact in the stay and performance equation. The other difference is that the positive correlation between the interviewer offer and stay effects becomes borderline significant at the 5% significance level. The overall similarity between the estimates in Table 6 and Table 4 suggests that attracting workers who tend to stay does not come at the expense of the quality of the workforce.

7.2 Attrition

Not all interviewers remained employed for the whole duration of our sample period: 37 interviewers left the company. The majority of these people were regular call agents (21) or non-senior HR employees (7). In contrast, only two senior HR interviewers and seven team managers left the company. Anecdotal evidence suggests that performance as an interviewer was not an important consideration in the context of job separations. The primary business of call agents is making calls, and that of team managers is to supervise agents: interviewing was only an intermittent extra duty. A somewhat similar argument applies to those who work at the HR division: interviewing was only part of the duties of those working at HR, with scheduling, training, and compensation policy occupying much of their time.

Formally, any difference in the characteristics of the interviewers who stay and who move are fully accounted for by the interviewer effects. In this context, we find that the stayers and the movers differ only with respect to their average acceptance effects: the stayers have on average a bit larger positive effect on the probability of accepting a job offer and this difference is statistically significant.

A more serious concern is that the characteristics that determine the interviewer's separation decision are correlated with the errors (unobserved factors) in the offer, acceptance, and stay equations. We test this hypothesis formally by applying the Nijman and Verbeek (1992) test for non-random attrition. One of the advantages of their approach is that it allows us to test the impact of interviewer attrition on the hiring process without formally modelling interviewer separation decisions. To implement the test, we measure whether the shocks that affect interviewer separation decisions are independent of the errors in the offer, acceptance, and stay equations for the candidates interviewed by the corresponding interviewer. In this implementation, the Nijman and Verbeek (1992) test consists of including in the offer, acceptance and stay equations a dummy that indicates whether the interviewer leaves the company in the next 30 days. Under the null hypothesis of independence, the coefficients of these dummy variables are equal to zero. The estimated coefficient is reported in Table 7: it is not significant at 5% significance level, and the point estimates are close to zero. We also consider a specification in which we include dummies for interviewer separation in one, two, and three months. Again, we do not find evidence that the shocks that determine interviewer separations affect also hiring and employment outcomes of the corresponding interviewees. Finally, we obtain similar results in our final specification in which we include dummies for interviewer separation in one, two, three, four, five and six months.

7.3 Interviewer Effects and Interviewee Characteristics

The final robustness exercise that we conduct is related to the relationship between the interviewer effects and the characteristics of the candidates. Table 3 provides evidence that the company does not employ any particular pattern to assign candidates to interviewers based on their observable characteristics. Table 8 extends this to outcome-relevant unobservable effects by showing that the correlations between the candidate characteristics and the estimated interviewer effects are essentially zero. In this sense, it is consistent with our findings based on Table 3 and confirms what we know about the hiring process in the company.

8 Comparative Statics

In this section we investigate the relative importance of the time-varying vs. time-invariant effects of interviewers, as well as the overall impact of interview-level heterogeneity, on offer, accept, and stay decisions. To do so, we perform comparative static exercises with the help of counterfactual experiments, thereby taking full advantage of the joint estimation of the offer, accept and stay equations.

8.1 Time-Varying Effects vs. Time-Invariant Effects

To compare the two effects, we investigate how the time-invariant and the time-varying interviewer effects compound with each other to create an overall interviewer effect. Charts 1.1-1.3 of Figure 5 plot the distributions of offer, accept, and stay probabilities for a candidate with the average characteristics across interviewers with and without experience. The results show that relative to the distribution of the offer probability for interviewers without any experience, interviewer experience boosts the probability of an offer across all interviewers. We document a similar qualitative effect of interviewer experience on the probability that a candidate accepts a job offer. Most importantly, we note that the magnitude of the effect of interviewer experience on the accept probability is actually larger than the impact of experience on the offer probability for the average candidate. In contrast, we see that interviewer experience has only a small positive effect on the stay probability for the average candidate. The plots also show that selectivity of candidates at the offer and accept stages has a relatively small effect on the heterogeneity in the hiring and employment outcomes for the average job candidate. Thus, the results show that both the time-varying and time-invariant interviewer effects have an economically significant and approximately equal in magnitude effect on the hiring outcomes for the average job candidate.

8.2 Hiring and Employment Outcomes

Estimating our model of the hiring process allows us to explore what would happen counter-factually if the company were able to exogenously alter the characteristics of its interviewers. In particular, since a company would likely wish to retain its best interviewers, we explore the effect of using only “good” interviewers versus only using “bad” interviewers. Since, as the bottom rows of Table 4 shows, interviewer quality can vary along multiple dimensions (an interviewer who makes more offers is not necessarily one who generates more acceptances even though they are correlated), there is some arbitrariness in ranking interviewers. For our measure, we have defined a “good” interviewer as one who is better than average on all three dimensions, and a “bad” interviewer as one who is worse than average on all three dimensions[6]. Charts 2.1-2.3 of Figure 5 simply show the distribution of outcomes for each group of interviewers. By construction, the "good" perform better than the "bad" on all three dimensions, though the effect is stronger for offers and stays than for acceptances. In the case of acceptances, the difference largely comes from a set of "good" interviewers with an acceptance probability in the range of 0.9-0.95.

Charts 3.1-3.3 of Figure 5 plot the distributions of offer, accept, and stay probabilities for the same two groups under a counterfactual exercise in which each set of interviewers is given a weighted, representative sample of all interviewees. The key difference with the actual outcomes is that all interviewers are now able to accumulate work experience equally, thus ignoring the effect of the actual work assignments implemented by the company. In other words, we evaluate the expected impact of interviewers on hiring and employment outcomes at the time of hiring behind a certain “veil of ignorance” about actual work assignments. As can be seen, this exercise accentuates

[6] This leaves out interviewers who are in the top half on at least one dimension and the bottom half in another.

the difference between the top and bottom interviewers in the context of the offer and acceptance decisions, but has fairly little additional effect on stays.[7]

Specifically, the charts suggest that the firm is not optimally using its best interviewers on any of the dimensions. Rather, the firm is actually offering more job experience to its worse interviewers (for example, those who have no call center experience, as discussed above), and therefore these worse interviewers are able to catch up somewhat with the better interviewers.

Table 10 quantifies the results in Figure 5. It compares the means and standard deviations associated with the offer, accept, and stay decisions for the "good" and for the "bad" in terms of actual (Actual) and expected performance under random-assignment (Random), along with expected performance under random assignment for those who work at HR but started elsewhere (Random+Exp.*). We show both the cumulative effects (e.g., the probability that an applicant will get an offer, accept the offer, and stay) as well as marginal effects (e.g., the probability that an applicant will stay given that she accepted an offer). The results from the table confirm the patterns detected in Figure 5. In all three scenarios, the "good" interviewers have both superior cumulative and marginal rates at each stage of the hiring process relative to the "bad." Moving from the actual to the counterfactual assignments, we note that the offer, accept and stay rates, both cumulative and marginal, for the "good" improve substantially. However, that is not the case for the"bad," where the cumulative and marginal acceptance rates actually slightly drop. We also formally test whether the offer, accept, and stay rates for the "good" interviewers under all three scenarios are statistically different from their counterparts for the "bad" interviewers under the actual assignment of interviews. In most (though not all) of the cases, the differences in the rates are statistically different

[7] Recall from Figure 2 that there is substantial evidence for an impact of interviewer experience on offers and acceptances, but fairly little cumulative learning of how to generate stays.

at the 1 percent significance level.

8.3 Hiring Costs

The preceding comparative statics led us to the conclusion that interviewer effects and experience have a large economic impact not only on the stay rates of employees but also on acceptance rates of job candidates. As a final exercise, we quantify the impact of interviewers on the firm's hiring costs. To monetize the effect of the differences in interviewer accept rates, we follow the methodology in Hoffman et al. (2016) with respect to recovering hiring costs per interviewer. In this setting, the strongest assumption that we make is that the hiring technology is linear with respect to the number of interviewed candidates, i.e. the company enjoys a constant marginal cost of hiring. To get a sense of the relative importance of interviewer heterogeneity, we also calculate the hiring costs as a percentage of annual expected operational profits of the company. Expected operational profits here are defined as the difference between the expected revenues and variable costs (total worker compensation) for an employee (expected profit from now on).

Table 11 reveals that the average hiring cost of filling in a vacant work station in the company is \$210, which amounts to about 4% of the expected profit from a newly hired employee. However, the cost for the "good" interviewers is actually \$191 (90% of the average cost or 3.6% of expected profit), while that for the "bad" interviewers is \$230 (109% of the average cost or 4.3% of expected profit). The difference between the hiring costs for the "good" interviewers under the actual and under random assignment is \$57 (27% of the average cost, which amounts to 1.07% of expected profit), while the difference between the hiring costs for the "good" interviewers under the actual assignment and our third assignment scenario amounts to \$69 (33% of the average cost or 1.3% of expected profit). Regardless whether these counter-factual actions are feasible in practice, we learn that one standard deviation in the hiring cost across interviewers amounts to 20-25% of the average hiring cost under the different scenarios (about 1% of expected profit). In other words, for every one hundred workstations that are filled-in by newly-hired employees, the difference between the hiring costs for the "good" and "bad" interviewers translates into gains equal to the operational profit from one extra employee.

In our setting, the call center company enjoys relatively low hiring costs in comparison with the call center companies in the sample of Hoffman et al. (2016). Even in such a setting, we document that the heterogeneity in the interviewer accept effects has a substantial impact on its hiring costs.

9 Discussion and Conclusions

This paper provides evidence that there is significant heterogeneity in the performance of interviewers. We find that heterogeneity exists along three dimensions of performance: filling seats, selling the job to prospective candidates, and screening for durable and productive hires. We find the strongest evidence of heterogeneity in interviewers' ability to sell the job, and the weakest in their ability to screen. Little of the heterogeneity can be explained by observable interviewer characteristics. Interviewers who are good at selling the job to candidates are also more likely to make offers. Moreover, this tendency does not come at the cost of screening for high-quality candidates. We show that interviewers' abilities to fill seats and sell the job improve with job tenure, while their ability to screen remains relatively constant over tenure. The impact of job tenure depends on the work history of the interviewers.

Our results have implications for understanding how interviewers add value during the job search process. The existence of persistent heterogeneity in interviewer performance leads us to conclude that interviewers are effective at both filling positions with high quality candidates and selling the company. Our findings complement recent literature that investigates the use of algorithmic technologies in the hiring process. We identify generating offers and acceptances as one of the most important roles human interviewers play – this of course is not what the current crop of technologies has been designed to do. Moreover, we also find evidence that in this specific context the interview process does in fact generate additional information and value on top of a pre-screening algorithm. Overall, the productivity advantage due to interviewer quality is substantial: If the company could conduct all of its interviews with its best interviewers, it would see costs savings of over 40 percent of its hiring and recruitment budget, which translates to 2 percent of operational profits.

As noted, much of the variation in interviewer quality is not explainable by characteristics that we are able to observe. Nevertheless, because interviewer performance is persistent, the firm could learn to screen and utilize better interviewers based on their early performance. We also note that one aspect of interviewer performance could be due to the firm's internal policies, as communicated to us by its managers: Interviewers are made acutely aware when they have not filled enough seats, but information about post-hire performance and retention of their interviewees is not sent back to the interviewers; much less are the interviewers evaluated on this performance dimension. We therefore cannot be sure whether interviewers are inherently weaker at screening candidates for durability and performance than they are at generating offer acceptances, or whether the firm has merely never incented or provided the relevant feedback for them to be good at it. Similarly, given the available information, we cannot be sure whether interviewers are good at screening for candidates who are likely to accept an offer, or that they exert effort to sell the company to each candidate and then make offers to the candidates who have been convinced. We leave these issues to future research.

References

Allen, Concha, Stacey Schetzsle, Michael L. Mallin and Ellen Bolman Pullins. 2014. Intergenerational recruiting: the impact of sales job candidate perception of interviewer age. *American Journal of Business* 29(2): 146-163.

Cowgill, Bo. 2019. Bias and Productivity in Humans and Machines. Working paper.

Dana, Jason, Robyn Dawes and Nathanial Peterson. 2013. Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making* 8(5): 512-520.

Dipboye, Robert L., Barbara B. Gaugler, Theodore L. Hayes, and Debra Parker. 2001. The Validity of Unstructured Panel Interviews: More than Meets the Eye? *Journal of Business and Psychology* 16(1): 35-49.

Fific, Mario and Gerd Gigerenzer. 2014. Are two interviewers better than one? *Journal of Business Research* 67: 1771–1779.

Glaser, R., P. A. Schwartz, and J. C. Flanagan. 1958. The contribution of interview and situational performance procedures to the selection of supervisory personnel. *Journal of Applied Psychology* 42: 69-73.

Heath, Rachel. 2018. Why do firms hire using referrals? Evidence from Bangladeshi garment factories. *Journal of Political Economy* 126(4): 1691-1746.

Heckman, James, and Salvador Navarro. 2007. Dynamic discrete choice and dynamic treatment effects." *Journal of Econometrics* **136 (2)**: 341-396.

Hoffman, Mitchell, Lisa Kahn and Danielle Li. 2017. Discretion in Hiring. *Quarterly Journal of Economics*. 133(2): 765–800.

Huffcutt, Allen and David Woehr. 1999. Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior* 20: 549 - 560.

Kausel, Edgar, Satoris S. Culbertson and Hector P. Madrid. 2016. Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes* 137(C): 27-44.

Knight, Rebecca. 2015. How to Conduct an Effective Job Interview. *Harvard Business Review*, January 23, 2015.

Maddala, Gangadharrao. 1983. *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press.

Maurer, S. D., and T. W. Lee. 2000. Accuracy of the situational interview in rating multiple candidates. J*ournal of Business and Psychology* 15: 73-96.

McDaniel, M.A., D. L. Whetzel, F. L. Schmidt and S. D. Maurer. 1994. The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology* 79(4): 599–616.

McMurry, R. N. 1947. Validating the patterned interview. *Personnel* 23: 263-272.

Nijman, T. and M. Verbeek. 1992.Testing for Selectivity Bias in Panel Data Models. *International Economic Review* 33(3). 681-703

Roulin, Nicolas, Adrian Bangerter and Elise Yerly. 2011. The Uniqueness Effect in Selection Interviews. *Journal of Personnel Psychology* 10: 43-47.

Schmidt, F. L., R. D. Zimmerman. 2004. A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology* 89: 553–561.

Table 1: Hiring and Employment Outcomes

	Cum.		Marginal	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Interviewed	0.437	0.496	0.437	0.496
Offered	0.238	0.334	0.545	0.498
Accepted	0.155	0.249	0.652	0.476
Stayed>6mo.	0.048	0.142	0.307	0.461
Obs.,	129,636		129,636	

Note: The table reports the share of initial applicants who reach successive stages of the hiring process and who remain in the firm for at least 6 months.

Table 2: Summary Statistics for Interviewers

Variable	Mean	Std. Dev.
Interviewer Tenure (in months)	14.957	13.615
Interviews per Interviewer per Day	5.530	4.684
Interviewer: Female	0.796	0.403
Interviewer: Black	0.372	0.483
Interviewer: Senior	0.135	0.342
Interviewer: Started Senior	0.125	0.331
Interviewer: HR	0.367	0.482
Interviewer: Started HR	0.290	0.454
Interviews	56,651	
Interviewers	87	

Note: Sample includes data from all US-based call centers between August 2012 and August 2017. The sample includes data on individuals who have been hired at least one year before the last date for which data are available.

Table 3: Correlations between Characteristics of the Interviewers and the Applicants

	Candidate: Female	White	Black	Age	Only HS	Test
Interviewer:						
Female	0.039	-0.089	0.120	0.027	0.001	-0.025
White	0.015	-0.005	0.048	0.046	-0.027	-0.033
Black	-0.005	-0.069	-0.023	-0.072	0.079	-0.012
Age	0.000	-0.029	0.031	0.102	0.033	-0.075
Test	0.021	-0.018	0.021	0.009	-0.099	-0.013
Senior	0.002	0.041	-0.102	-0.041	0.001	0.092

Note: The table reports the correlations between observable characteristics of the candidates and their interviewers after controlling for location effect by subtracting the corresponding location averages.

Table 4: Estimates of Baseline Selection Model

	Offer		Accept		Stay>6mo.	
Variables:	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intwr. Ten.	-0.006	0.008	0.046**	0.011	0.008	0.022
Intwr. Ten.*(Start Sr.)	0.039**	0.015	-0.023	0.017	-0.015	0.021
Intwr. Ten.*(Became Sr.)	-0.001	0.010	0.000	0.010	0.021	0.012
Intwr. Ten.*(Start HR)	0.006	0.012	0.005	0.011	0.031	0.019
Intwr. Ten.*(Moved HR)	0.016	0.016	0.031*	0.016	0.076**	0.020
Prior Intws.	0.023**	0.004	-0.007	0.006	-0.004	0.010
Prior Intws.*(Start Sr.)	-0.016	0.155	0.008	0.101	-0.005	0.121
Prior Intws.*(Became Sr.)	0.181	0.155	-0.039	0.129	-0.220	0.148
Prior Intws.*(Start HR)	-0.034	0.110	0.005	0.127	-0.012	0.144
Prior Intws.*(Moved HR)	-0.198	0.160	-0.118	0.133	-0.290	0.155
Suggested: Jr. Agent	0.148**	0.028	0.063	0.037	0.111*	0.050
Suggested: Agent	0.250**	0.028	0.169**	0.037	0.142*	0.058
Suggested: Sr. Agent	0.123*	0.057	0.276**	0.073	0.378**	0.105
Same Ethnicity	0.014	0.019	-0.012	0.025	-0.028	0.034
Same Sex	-0.013	0.019	0.012	0.022	-0.013	0.029
Candidate: Younger	-0.062**	0.020	-0.148**	0.026	-0.058	0.046
Candidate: Older	0.106**	0.025	0.352**	0.033	0.051	0.080
Past Intws on Day	-0.008**	0.001	-0.004*	0.002	0.000	0.003
Days to Intw	-0.002*	0.001	0.031*	0.013		
Days to Offer	-0.002*	0.001	-0.032*	0.013		
Past Avg. Test Score	0.004*	0.002				
Const.	0.593**	0.179	1.683**	0.220	-1.634**	0.363
	$Corr(o,a)$		$Corr(o,s)$		$Corr(a,s)$	
	Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
Corr.: Errors	0.320**	0.113	0.403	0.209	0.174	0.378
Corr.: Intwr Effects	0.464**	0.096	0.180	0.106	-0.142	0.107

Note: The specification includes the following candidate characteristics: age, gender, ethnicity, educational attainment, general work experience, prior call center experience, education, entry test score, referral status, applicant cohort dummies, and location dummies. It also includes the interviewer evaluation of the applicants with respect to speaking and listening skills. In terms of unreported exclusion restrictions, the offer equation contains dummies for the day of the week on which the test is taken. In the correlations, o stands for offer, a for accept, s for stay>6mo. ** represent 1 percent significance level; * represents 5 percent significance level. Obs, for offer decision: 57,333; obs, for accept decision: 31,328; Obs, for stay decision: 16,415. Log-likelihood: -57817.689

Table 5: Tests for Significance of Interviewer Effects

Specification:	Log-likelihood	d.f.	Test stat.	Significance
No Intwr. Offer Effects	-59145.964	87	2656.55	0.001
No Intwr. Accept Effects	-58005.072	87	374.766	0.001
No Intwr. Stay Effects	-57930.317	87	225.256	0.001
Only Extreme Intwr. Stay Effects	-57885.287	65	135.196	0.001
No Intwr. Effects	-59460.219	261	3285.06	0.001
Unrestricted	-57817.689			

Note: The table includes the results from the likelihood ratio tests for the joint significance of the interviewer effects in each equation. Extreme interviewer stay effects are defined as the top and botom 5 percent of the interviewer stay effects under the unrestricted specification. The corresponding restricted specification sets all other interviewer stay effects to zero.

Table 6: Estimates of Selection Model, Including Performance

	Offer		Accept		Stay>6mo & Perf.>3	
Variables:	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intwr. Ten.	-0.006	0.008	0.046**	0.011	0.005	0.024
Intwr. Ten.*(Start Sr.)	0.039**	0.014	-0.023	0.015	-0.003	0.024
Intwr. Ten.*(Became Sr.)	-0.001	0.010	0.000	0.010	0.006	0.013
Intwr. Ten.*(Start HR)	0.022	0.014	0.006	0.013	-0.008	0.012
Intwr. Ten.*(Moved HR)	0.015	0.016	0.032*	0.015	0.049*	0.020
Prior Intws.	0.023**	0.004	-0.007	0.006	0.001	0.011
Prior Intws.*(Start Sr.)	-0.016	0.115	0.008	0.118	-0.005	0.162
Prior Intws.*(Became Sr.)	0.181	0.155	-0.41	0.128	0.314	0.171
Prior Intws.*(Start HR)	-0.034	0.110	0.002	0.111	0.006*	0.108
Prior Intws.*(Moved HR)	0.164	0.160	0.122	0.133	0.425*	0.179
Suggested: Jr. Agent	0.147**	0.028	0.063	0.037	0.111*	0.050
Suggested: Agent	0.249**	0.028	0.169**	0.037	0.142*	0.058
Suggested: Sr. Agent	0.122*	0.057	0.276**	0.073	0.378**	0.105
Same Ethnicity	0.014	0.019	-0.013	0.024	-0.028	0.034
Same Sex	-0.014	0.019	0.012	0.022	-0.013	0.029
Candidate: Younger	-0.062**	0.020	-0.147**	0.026	-0.058	0.046
Candidate: Older	0.107**	0.025	0.352**	0.033	0.051	0.080
Past Intws on Day	-0.008**	0.001	-0.004*	0.002	0.000	0.003
Days to Intw	0.002*	0.001	0.031**	0.013		
Days to Offer	-0.002*	0.001	-0.032**	0.013		
Past Avg. Test Score	0.003	0.003				
Const.	0.599**	0.124	1.677**	0.219	-1.735**	0.389
	$Corr(o,a)$		$Corr(o,s\&y)$		$Corr(a,s\&y)$	
	Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
Corr.: Errors	0.326**	0.111	0.383	0.252	-0.010	0.412
Corr.: Intwr Effects	0.305**	0.103	0.274*	0.104	-0.113	0.107

Note: The specification includes the following candidate characteristics: age, gender, ethnicity, educational attainment, general work experience, prior call center experience, education, entry test score, referral status, applicant cohort dummies, and location dummies. It also includes the interviewer evaluation of the applicants with respect to speaking and listening skills. In terms of unreported exclusion restrictions, the offer equation contains dummies for the day of the week on which the test is taken. In the correlations, o stands for offer, a for accept, s and y for stay>6mo and performance>3. ** represent 1 percent significance level; * represents 5 percent significance level. Obs, for offer decision: 57,333; obs, for accept decision: 31,328; Obs, for stay decision:16,415. Log-likelihood: -56673.588

Table 7: Test for Non-Random Interviewer Attrition

	Offer		Accept		Stay >6mo.	
Variables:	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Test 1:						
Intwr Separation in 1 mo.	0.015	0.043	-0.029	0.061	-0.026	0.078
Test 2:						
Intwr Separation in 1 mo.	0.017	0.051	-0.076	0.070	-0.040	0.090
Intwr Separation in 2 mo.	-0.008	0.048	-0.082	0.065	0.008	0.085
Intwr Separation in 3 mo.	0.028	0.045	-0.054	0.062	-0.063	0.078
Test 3:						
Intwr Separation in 1 mo.	0.069	0.059	-0.105	0.078	-0.065	0.104
Intwr Separation in 2 mo.	0.043	0.056	-0.110	0.073	-0.025	0.098
Intwr Separation in 3 mo.	0.079	0.055	-0.082	0.071	-0.090	0.093
Intwr Separation in 4 mo.	0.000	0.052	-0.111	0.070	-0.065	0.092
Intwr Separation in 5 mo.	0.014	0.052	-0.001	0.063	0.005	0.083
Intwr Separation in 6 mo.	0.017	0.051	0.022	0.061	-0.048	0.083

Note: The test specification follows Nijman and Verbeek (1992) and includes indicators whether the interviewer quit the same month, the following month, etc. It also includes the interviewer and interviewee variables of the baseline model. ** represent 1 percent significance level; * represents 5 percent significance level. Obs, for offer decision: 57,333; obs, for accept decision: 31,328; Obs, for stay decision: 16,415.

Table 8: Correlations between Characteristics of the Applicants and Interviewer Effects

	Interviewer: Offer Effect	Accept Effect	Stay effect
Candidate:			
Test	0.021	-0.048	0.055
Only HS	-0.039	-0.042	0.034
Female	-0.048	0.032	-0.035
White	0.103	0.001	0.041
Black	-0.123	0.059	-0.059
Age	0.015	0.100	-0.051

Note: The table reports the correlations between observable characteristics of the candidates and the estimated offer, acceptance, and stay interviewer effects in the baseline model.

Table 9: Regression Analysis of Interviewer Effects

Variables	$O.E.$	$A.E.$	$S.E.$	$O.E.$ resid
Black	0.118	0.328*	0.004	0.004
	(0.213)	(0.131)	(0.211)	(0.182)
Female	-0.886*	0.121	-0.412	-0.731*
	(0.244)	(0.151)	(0.239)	(0.206)
Test Score	0.003	0.015	0.015	-0.012
	(0.010)	(0.006)	(0.010)	(0.009)
Senior	0.887*	0.333	-0.014	0.751*
	(0.325)	(0.200)	(0.316)	(0.272)
Started Sr.	-0.787	-0.128	0.057	-0.782
	(0.525)	(0.323)	(0.524)	(0.452)
HR	0.072	0.388*	-0.350	-0.357
	(0.224)	(0.138)	(0,222)	(0.340)
Started HR	-0.306	0.216	-0.214	0.130
	(0.343)	(0.212)	(0,394)	(0.192)
R- Sqr	0.22	0.30	0.11	0.3
Obs.	87	87	87	87

Note: The table reports the regression of offer (O.E.), accept (A.E.), stay (S.E.), and residual offer (O.E. resid) effects. The offer effect residuals come from a regression of the interviewer offer effects on the interviewer accept and stay effects, in which the R-squared is 0.37. The constant terms in the regressions are not reported. Std. dev. are in parentheses, **=1 percent significance level, and *=5 percent significance. level.

Table 10: Comparative Statics on the Heterogeneity of Interviewer Effects

Outcomes	Bad						Good					
	Actual		Random		Random+Exp.*		Actual		Random		Random+Exp.*	
	Cum.	Marg.	Cum.	Marg.	Cum.	Marg.	Cum.	Marg.	Cum.	Marg.	Cum.	Marg.
Offers	0.514	0.514	0.512	0.512	0.352^{**}	0.352^{**}	0.641^{**}	0.641^{**}	0.669^{**}	0.669^{**}	0.787^{**}	0.787^{**}
	(0.499)	(0.499)	(0.495)	(0.495)	(0.478)	(0.478)	(0.479)	(0.479)	(0.396)	(0.393)	(0.409)	(0.409)
Hired	0.359	0.698	0.334^{**}	0.652^{**}	0.254^{**}	0.721^{**}	0.424^{**}	0.663^{**}	0.471^{**}	0.704	0.636^{**}	0.808^{**}
	(0.480)	(0.459)	(0.471)	(0.476)	(0.435)	(0.449)	(0.500)	(0.473)	(0.445)	(0.432)	(0.481)	(0.399)
Stay>6mo	0.094	0.263	0.162^{**}	0.486^{**}	0.182^{**}	0.717^{**}	0.159^{**}	0.374^{**}	0.293^{**}	0.623^{**}	0.474^{**}	0.745^{**}
	(0.282)	(0.440)	(0.362)	(0.440)	(0.385)	(0.451)	(0.358)	(0.484)	(0.377)	(0.406)	(0.499)	(0.436)

Note: The table presents a comparison between the realized performance (Actual) of Good and Bad interviewers. Good interviewers (Good) are defined as people who have better than average interviewer effects on all three dimensions, and bad interviewers (Bad) are defined as people who have worse than average interviewer effects on all three dimensions. It also includes the results from a counterfactual experiment under which interviewers get equal opportunity to accumulate experience (Random) and from a modification of the random assignment in which all interviewers start working elsewhere in the company before joining the HR division (Random+Exp.*). The table reports the cumulative (Cum.) and marginal (Marg.) probability rates and below them the std. dev. in parentheses. **= the hypothesis that the actual rates for the Bottom group are equal to their counterparts in the other scenarios is rejected at the 1 percent significance level.

Table 11: Impact of Interviewer Accept Effects on Hiring Costs

Outcomes (in $):	Mean ($)	Std. Dev.($)	% of E(Profit)
Hiring costs per workstation: Actual	209.65	45.83	3.95
Hiring costs per workstation, Bottom: Actual	229.76	36.74	4.31
Hiring costs per workstation, Top: Actual	190.93	54.92	3.61
Hiring costs per workstation, Top: Random	134.18	60.81	2.53
Hiring costs per workstation, Top: Random+Exp.*	121.87	59.87	2.30

Note: Hiring costs per workstation are defined as the expected costs of filling in a vacant workstation. Good interviewers (Good) are defined as people who have better than average interviewer effects on all three dimensions. Bad interviewers (Bad) are defined as people who have worse than average interviewer effects on all three dimensions. We compare three types of interviewer assignment: the actually implemented assignment (Actual); random assignment under which interviewers get equal opportunity to accumulate experience (Random), and a modification of the second assignment in which all interviewers start working elsewhere in the company before joining the HR division.

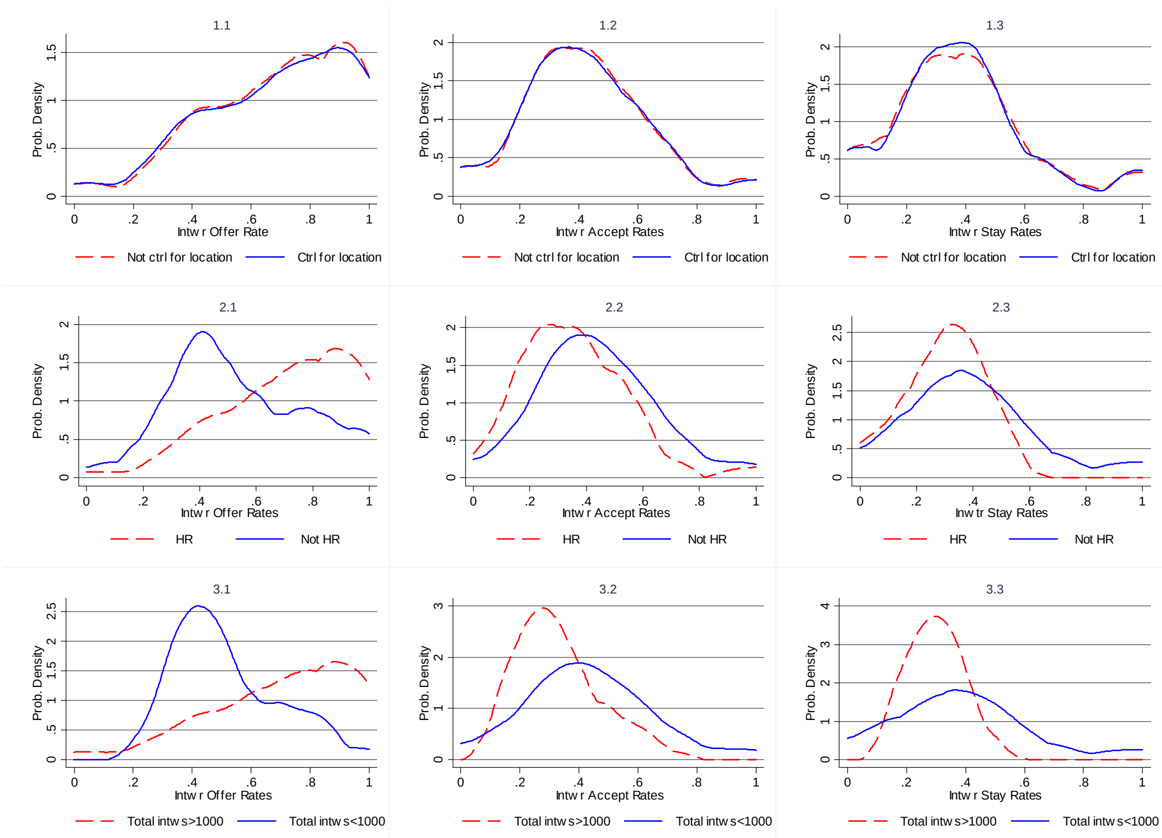

Figure 1: Heterogeneity in the impact of interviewers on hiring and employment outcome. Charts 1.1-1.3 plot the distributions of interviewer offer, accept, and stay rates. Charts 2.1-2.3 plot the distributions of interviewer offer, accept, and stay rates, controlling for location effects. Charts 3.1-3.3 plot the distributions of interviewer offer, accept, and stay rates for interviewers with more than 500 interviews, controlling for location effects.

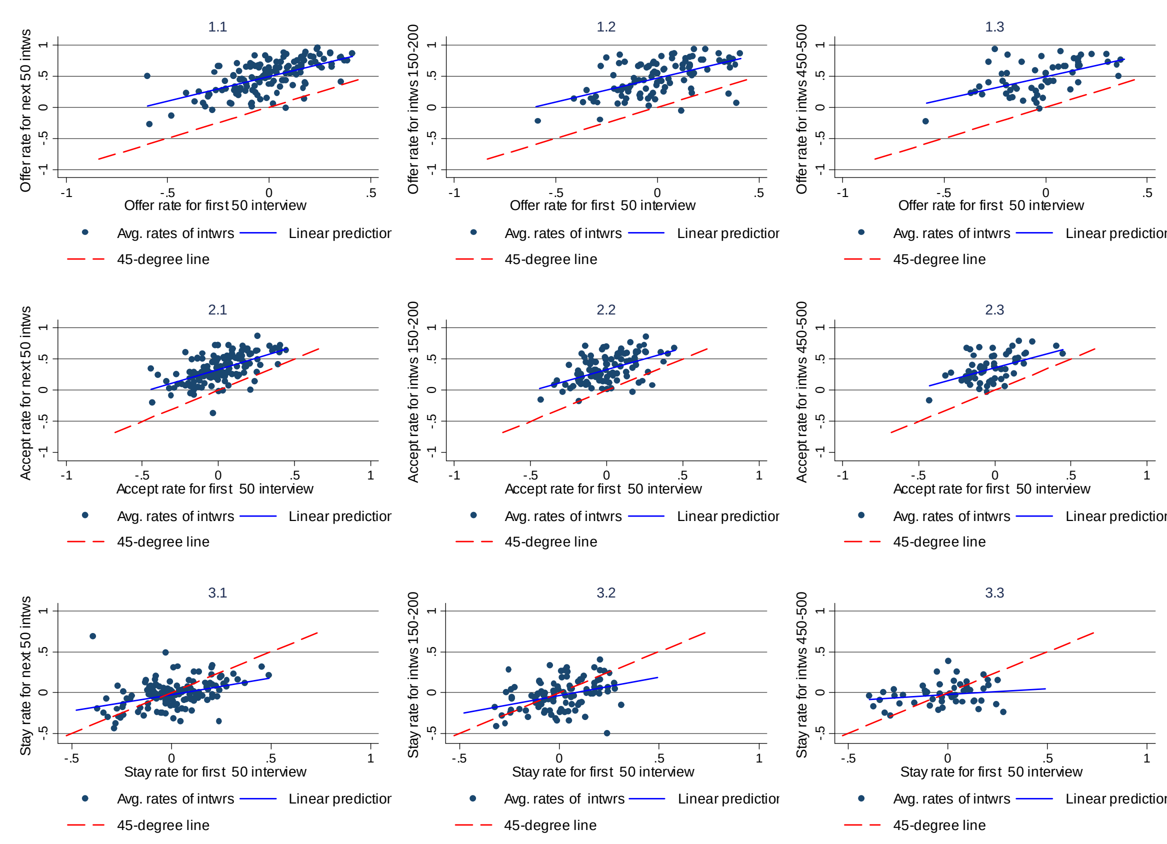

Figure 2: Relationship between past and future offer, acceptance, and stay rates of interviewers. Charts 1.1-1.3 explore whether the interviewer offer rate for the first 50 interviews predicts the interviewer offer rates for interviews 51-100, 151-200, 451-500. Charts 2.1-2.3 test whether the interviewer accept rate for the first 50 interviews predicts the interviewer accept rates for interviews 51-100, 151-200, 451-500. Finally, charts 3.1-3.3 investigate whether the interviewer stay rate for the first 50 interviews predicts the interviewer stay rates for interviews 51-100, 151-200, 451-500.

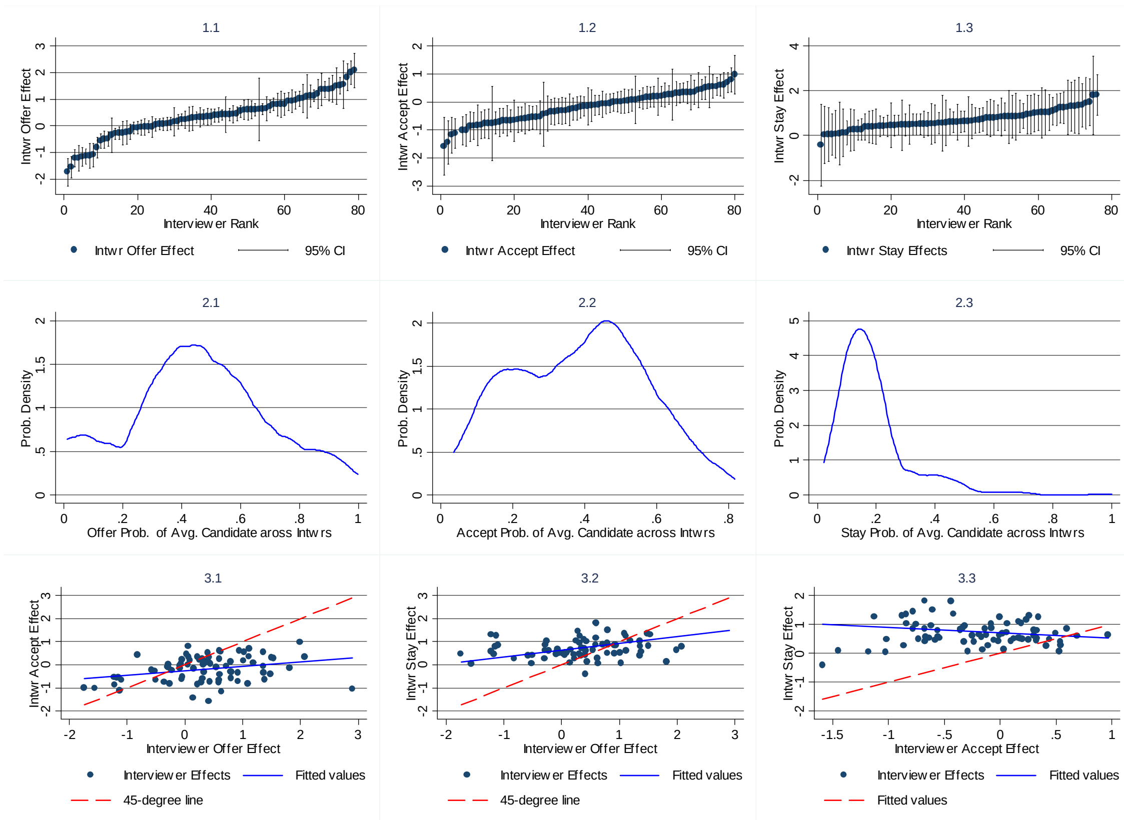

Figure 3: Estimated interviewer offer, acceptance, and stay effects. Charts 1.1-1.3 plot the point estimates, along with the corresponding 95-percent confidence intervals. Charts 2.1-2.3 plot the distributions of the interviewer offer, accept, stay effects, along with the corresponding distributions of the interviewee related characteristics. Charts 3.1-3.3 plot the correlations between the interviewer effects.

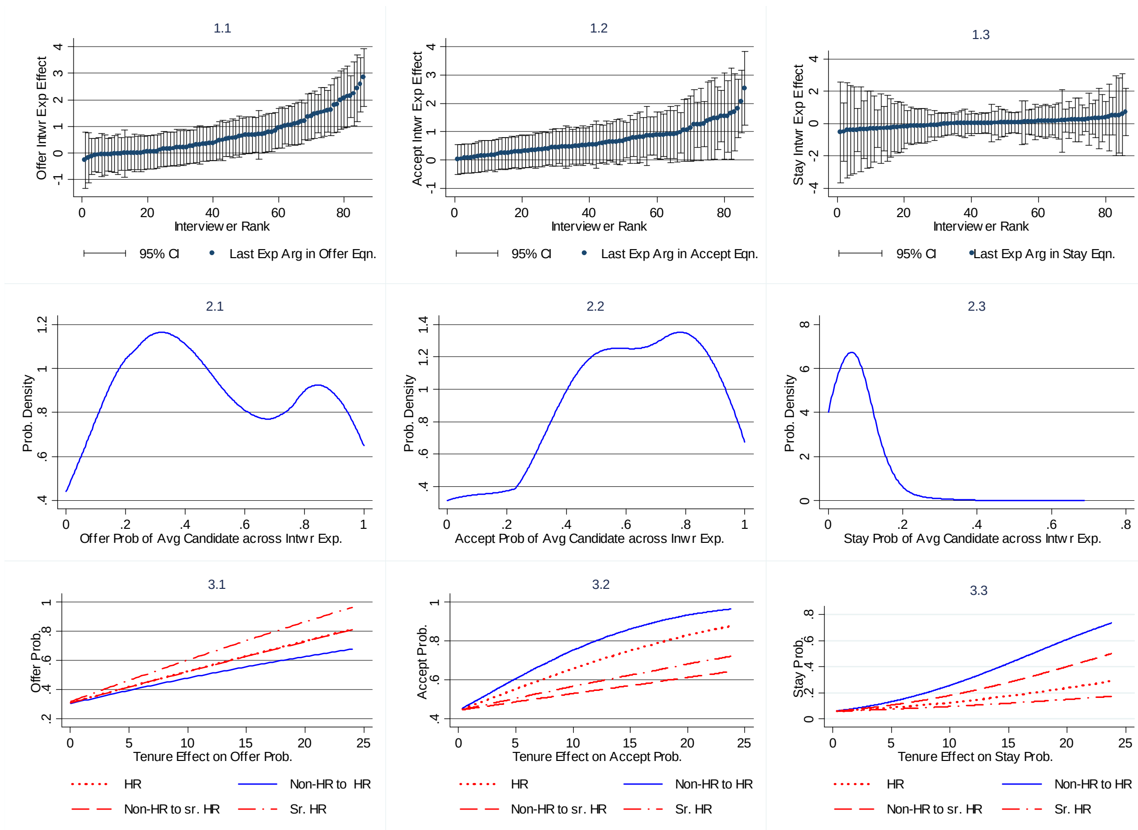

Figure 4: Estimated effect of interviewer experience on offer, acceptance, and stay decisions. Estimated effect of interviewer experience on offer, acceptance, and stay decisions. Charts 1.1-1.3 plot the point estimates of the interviewer transient, or time-varying, effect on the last recorded offer, accept, and stay decisions, along with the corresponding 95-percent confidence intervals. Charts 2.1-2.3 plot the distributions of the interviewer experience effects, along with the corresponding distributions of the interviewee related characteristics. Charts 3.1-3.3 plot the effect of tenure on the experience terms in the offer, accept, and stay equations.

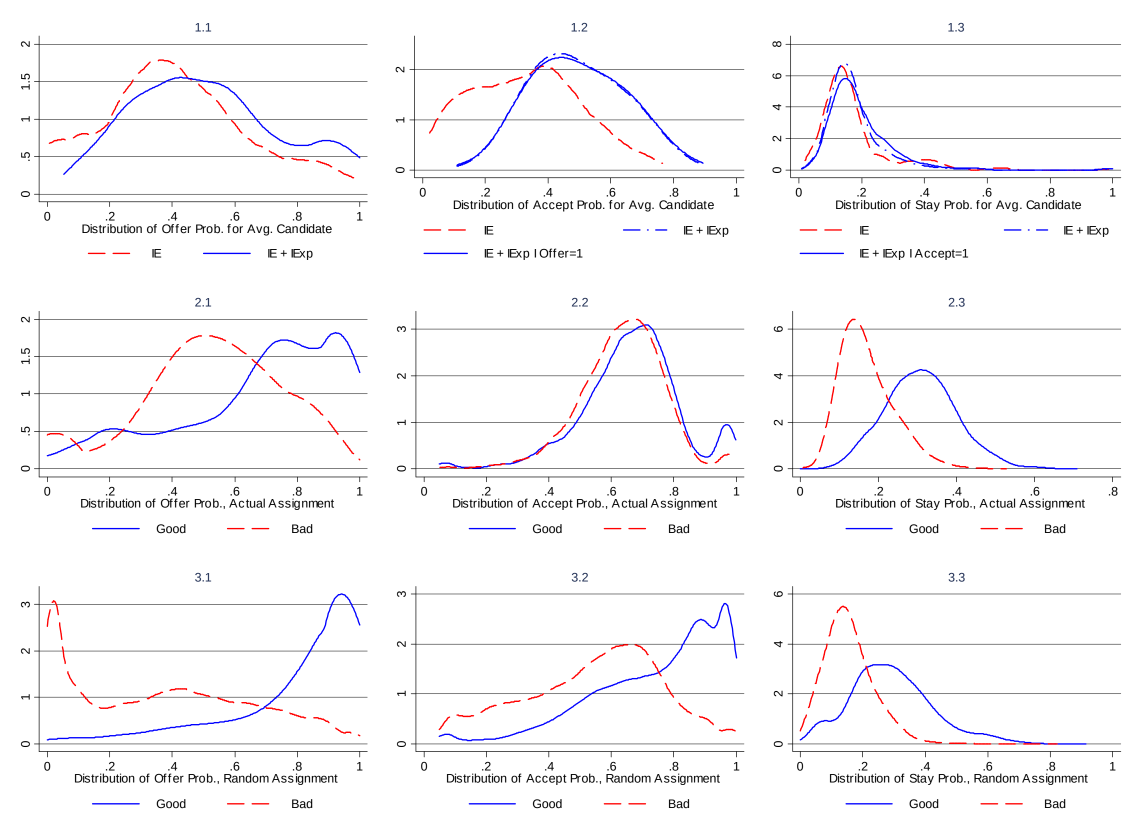

Figure 5: Comparative statics. We consider two groups of interviewers: those with effects in the top two quartiles of the distributions of interviewer offer, accept, and stay effects (Good) and those with effects in the bottom two quartiles of the distributions of interviewer offer, accept, and stay effects (Bad). Charts 1.1-1.3 plot the distributions of offer, accept, and stay probabilities for a candidate with the average characteristics across interviewers with (I.E.+ I.Exp) and without experience (I.E.). Charts 2.1-2.3 plot the distributions of actual offer, accept, and stay probabilities for the Good and Bad groups. Charts 3.1-3.3 plot the distributions of expected offer, accept, and stay probabilities for the same two groups.