Should Human Capital Development Programs be Mandatory or Voluntary? Evidence from a Field Experiment on Mentorship

Jason Sandvik* Richard Saouma † Nathan Seegert‡

Christopher Stanton§

Abstract

In a field experiment, we find that a mandatory mentorship program raises worker productivity while a voluntary version of the program does not. A significant reason why the mandatory program results in larger gains is that the lowest productivity employees do not participate when the program is voluntary, despite their having the greatest treatment benefits. A nationally representative survey of U.S. workers shows wide variation in human capital development program participation, suggesting that understanding self-selection is important for firms' returns on these programs across a variety of settings. Our findings have implications for resource allocation, experimental design, productivity dispersion, and inequality.

JEL: C93 M53 M54
Keywords: Human Capital Development, RCT, Mentorship, Training, Personnel Economics, Program Evaluation

We thank Emily Beam, Jasmijn Bol, Laura Boudreau (discussant), Zoe Cullen, Florian Englmaier (discussant), Guido Friebel, Robert Garlick, Robert Gary-Bobo (discussant), Isaac Hacamo (discussant), Jessica Hoel, Mitch Hoffman, Lisa LaViers, John List, Michelle Lowry, Bentley MacLeod, Robert Metcalfe, Paige Ouimet, Dimitris Papanikolaou, Raffaella Sadun, Elena Simintzi, Jason Snyder, Harish Sujan, Brian Waters (discussant), Michael Weisbach, seminar participants at Harvard Business School, MIT, the University of Michigan's Ross Strategy Brownbag series, the Indian Institute of Management Ahmedabad, the University of Arizona, and conference participants at the CEPR IMO, the Econometric Society Latin American Meeting, the FMA, the 2022 Labor and Finance Meeting, the Advances in Field Experiments Meeting, the 2022 Strategy Science Conference, the 2022 NBER Summer Institute, the 2022 NBER Organizational Economics Meeting, and the 2023 AEA Annual Meeting for helpful comments.

*Eller College of Management, University of Arizona
†Eli Broad College of Business, Michigan State University
‡David Eccles School of Business, University of Utah
§Harvard Business School, NBER, and CEPR

1 Introduction

Organizations actively train, educate, and upskill their employees through human capital development programs. These programs represent an increasingly large component of corporate spending and are a growing driver of firm value (Zingales, 2000; Edmans, 2011; Bloom et al., 2016; Nishesh et al., 2022; Rouen and Regier, 2022). While U.S. firms' training investments alone now total over $100 billion annually (Statista, 2022), relatively little research addresses how organizations choose to allocate development resources across workers.[1] We conduct a novel field experiment to estimate the trade-offs involved around a ubiquitous allocation decision: whether to make employee development programs mandatory or voluntary. Whether a mandatory or voluntary program is more effective—and improves resource allocations—hinges on workers' self-selection. If those who benefit most opt into voluntary programs, then self-selection enables firms to target resources to employees with the highest return. If, instead, workers volunteer at random (or worse, if participation is negatively correlated with program gains), then mandates may be needed to efficiently allocate resources among workers.

We investigate the effectiveness of a mentorship program and the implications of making it mandatory or voluntary in a field experiment involving a large U.S.-based sales organization.[2] Salespeople at the firm answer incoming calls to sell digital subscriptions (e.g., television, internet, and cellular services), and their incentives are likely aligned with the firm to increase their own human capital (Zivin et al., 2021), as commissions comprise over a third of the median employee's compensation.[3] This setting is well-suited to evaluate the design of human capital development programs for at least five reasons: (i) sales agents work independently of each other, (ii) we have individual, daily sales performance data for

[1]The prior literature on human capital development has largely focused on whether firms can rationalize training investments (Becker, 1975; Acemoglu and Pischke, 1999; Fudenberg and Rayo, 2019; Starr, 2019), rather than how program features influence returns.

[2]Workplace mentorship programs are themselves ubiquitous in the U.S., with approximately 70% of the Fortune 500 firms offering such programs (Gutner, 2009).

[3]Commission rates rise with performance, with the most (least) productive workers earning commissions equal to 8% (3%) of their total sales revenue each week.

all workers, (iii) inbound calls are randomly assigned to sales agents, (iv) the firm regularly hires new sales agents in cohorts that train together, allowing the program to be administered under different conditions for similar groups, and (v) there is variation in productivity across new hires (e.g., new agents at the 75th percentile of the sales revenue distribution generate twice as much revenue as those at the 25th percentile), allowing us to characterize differences in who selects into the program and who benefits most from it across the productivity distribution. Although these features are attractive for estimation, the evidence likely generalizes to many other work settings. There are over four million jobs in sales and customer service occupations in the United States where workers do similar tasks to those at the study firm (American Community Survey, 2019). In addition, a nationally representative survey that we fielded confirms that: a) the program we test exists in many companies, b) companies vary in whether they make their programs mandatory or voluntary, and c) many workers choose not to participate in voluntary programs.

Our experiment entailed two levels of randomization: the first (high level) is at the cohort level for training classes of new hires, and the second (low level) is at the worker level within a cohort. At the high level, new hire cohorts were randomized into one of two groups, labeled the Mandatory-Condition and the Voluntary-Condition. In Mandatory-Condition cohorts, the lower-level treatment involved randomly assigning agents to have a mentor or not, i.e., agents were not first asked whether or not they wanted to participate in the program. For Voluntary-Condition cohorts, on the first day of training, the firm's staff briefly described the mentorship program and asked each new hire to privately indicate whether they had an interest in participating in the program. We label those who indicated an interest in participating as, "opted in"; for those who opted in, the lower-level treatment involved assigning a random subset to receive a mentor. Agents who "opted out" of participation were not assigned a mentor. All mentor assignments were randomly drawn from a pool of established sales agents who had no formal authority over those they mentored (i.e., mentors were more experienced, lateral peers without supervisory responsibilities). Mentors volunteered to be in the program with the understanding that participation would look favorable for future promotions. Matched mentor-protégé pairs were asked to meet for 30 minutes each week for four weeks and to record what they discussed on a worksheet. Mentors were not informed of which condition, mandatory or voluntary, their matched protégé came from.

We find that mentorship had a positive and statistically significant effect on workers' productivity in the Mandatory-Condition. Specifically, workers who received a mentor increased their individual output by about 0.145 standard deviations when using a combined index of four productivity measures. For mentored workers, daily sales revenue and revenue-per-call (the firm's two focal performance measures) increased by 19% and 12%, respectively, compared to non-mentored agents over their first two months of tenure. The primary mechanism appears to be knowledge transfer, as about 45% of the treatment effects persist through agents' first six months of tenure, well after the program's conclusion. A follow-up survey indicated that the program allowed treated agents to ask questions and receive help, consistent with improved psychological safety and subsequent knowledge transmission (Edmondson and Lei, 2014; Chandrasekhar et al., 2018; Castro et al., 2022). In addition, worksheets that mentor-protégé pairs completed during the program indicate that knowledge exchange occurred.

Next, we test whether firms can enhance average program outcomes by making the program voluntary. If workers who benefit most participate when the mentorship program is optional, then a voluntary program allows the firm to target those with the highest gains. Alternatively, a voluntary program could have no difference in effectiveness if treatment effects are uniform across workers, or it could even dampen the benefits if there are heterogeneous treatment effects and self-selection is negatively correlated with treatment gains. We may also find different effects in the voluntary program for other reasons, such as program framing. For example, workers could be more engaged with the program if they were the ones to volunteer for it, or they could be less engaged if the program's voluntary nature signals a lack of importance.

We find that mentorship did *not* affect the productivity of workers who opted into the program in the Voluntary-Condition. Specifically, those who opted into the program and were randomly assigned a mentor had similar levels of productivity to those who opted in and were not assigned a mentor.

To assess why treatment affects differ between the Mandatory- and Voluntary-Conditions, we next quantify the relative importance of self-selection, heterogeneous treatment effects, and other mechanisms, such as program framing. We quantify the degree of self-selection as a function of realized productivity by comparing sales output for those who opt into the program and are not assigned a mentor, with those who opt out. Workers who opt out are much less productive, with revenue between 23 and 30 percent lower than untreated workers who opt into the program. The strongest predictor of opting out is a pre-hire assessment score, given by interviewers during the recruitment process. Workers with low pre-hire assessments are much more likely than others to opt out of program participation. In contrast, workers' demographics, work history, and personality characteristics (collected via surveys) have little predictive power for the opt-out decision.[4] We use these factors to predict a propensity score for whether agents in the Mandatory-Condition would have been more or less likely to opt out of the program had they been given the choice.

When we quantify heterogeneous treatment gains in the mandatory program based on agents' propensity to opt out of mentorship, we find that workers who are the least likely to

[4]We do not evaluate the extent to which workers' beliefs about the efficacy of the program predicted their opt-out decisions, as eliciting subjects' ex-ante beliefs about treatment effects could have potentially swayed their participation decision. To get at belief-related mechanisms behind the opt-out decision, we rely on evidence from a nationally representative survey of workers. In that survey, we find wide variation in firms' practices regarding whether programs are mandatory or voluntary, with substantial rates of non-participation in voluntary programs. Workers cite time constraints and inconvenience as the primary reasons for non-participation; these issues are less applicable to the experiment, as the program was conducted during work hours. The next most common reasons highlight skepticism about personal benefits and the desire to avoid interacting with coworkers or bosses. Intimidation around interacting with more productive coworkers was a theme that emerged in interviews during our prior work in this firm (Sandvik et al., 2020), but to the extent that variation in personality characteristics may pick up differential propensity for intimidation, we do not detect evidence for it. A related possibility is that some workers may believe program enrollment signals something about their competence, which has been shown to impact advice seeking at work (Heursen et al., 2023).

participate in the program benefit most. We estimate separate treatment effects for agents in the top tercile of the opt-out propensity score distribution (i.e., those with a high likelihood of opting out) and for those in the bottom two terciles (i.e., those with a low likelihood of opting out).[5] The mentorship treatment significantly raised the productivity of those agents in the top tercile (labeled likely-to-opt-out agents), whereas the effect of mentorship was significantly weaker among agents who were less likely to opt out. Furthermore, likely-to-opt-out agents generate significantly less revenue than those in the bottom two terciles when untreated, consistent with the previously discussed selection effects. This exercise suggests that self-selection and heterogeneous treatment effects explain about one-third of the overall gap in treatment effects between conditions.[6]

The remaining difference in treatment effects between the Mandatory- and Voluntary-Conditions is due to other factors, such as program framing. Program framing can impact overall buy-in and engagement, including that from workers who would have participated in the voluntary program were they proffered the choice. We find evidence suggesting that framing is likely important, as treated agents in the Mandatory-Condition who were likely to participate in the program had greater sales gains due to mentorship than similar workers in the Voluntary-Condition. In addition, treated agents in the Mandatory-Condition were more likely to meet with their mentors than were treated agents in the Voluntary-Condition. Although we note that the program framing explanation is indirect and based on suggestive evidence, other channels are less likely to explain the difference in treatment effects across conditions. For instance, in both conditions, the mentorship program had no impact on retention, and three empirical exercises suggest that retention differences do not explain the sales revenue treatment effects. Furthermore, we designed the experiment to test for information leakages, other violations of the Stable Unit Treatment Value Assumption (SUTVA),

[5]We use terciles because the propensity to opt out is noisily estimated and we lose power when using finer partitions of the sample.

[6]The gains in the Mandatory-Condition for workers who are likely to be lower-performers suggest that any incentive conflict between the firm and these workers (who have relatively lower anticipated commission rates than higher-performers) can be overcome with the firm's guidance to use program resources for improvement.

and crowding out of organic mentorship. We find no evidence for these channels.

The firm realized significant benefits from implementing the mandatory version of the mentorship program. To quantify the gains, we calculate the return on investment across all mentoring relationships over a six-month horizon from treatment. We account for worker turnover by filling in a random, non-mentored replacement agent when we observe a worker leave the firm (either treated workers or those in the control group). Using this approach, we find that the firm gained $536,000 in revenues from treating 127 agents in the Mandatory-Condition over a six-month post-treatment horizon. The total costs, including overhead costs, associated with the mandatory program were $97,000. As such, the firm realized a $453,000 return on a $97,000 investment by implementing the mandatory mentorship program.

The program implementation had substantial implications for returns to the firm. Had the voluntary program instead been mandatory, the firm would have gained an additional $207,000, assuming one-third of the treatment effect difference due to self-selection and heterogeneous effects carries over from the Mandatory-Condition. Although our results suggest firms cannot always rely on self-selection to allocate workers with the highest returns into human capital development programs, in some settings alternative allocation approaches may be superior to mandatory rules that draw in all workers (Li et al., 2020; Johnson et al., 2023). In those cases, testing absent selection concerns using the approach in the Mandatory-Condition can inform refinements to resource allocation decisions. In practice, firms face a trade-off between gathering data for better targeting up-front and the horizon over which treatment gains are realized. In workplaces with high rates of attrition, the delay costs of gathering data may outweigh the misallocation costs of treating too many workers.

Our findings speak to the efficacy of human capital development programs, who benefits from these programs, how firms choose to deploy them, and how data from pilots that evaluate them should be interpreted. First, we provide evidence on the effectiveness of mentorship programs, advancing the understanding of whether widespread adoption of mentoring is justified (Hilmer and Hilmer, 2007; Gutner, 2009; Lyle and Smith, 2014; Ginther et al., 2020). Addressing this question inside firms has been challenging due to nonrandom selection in most mentorship settings (Allen et al., 2017). Related work has studied the efficacy of other types of workplace programs, like purpose workshops and wellness programs (Ashraf et al., 2024; Gubler et al., 2018; Jones et al., 2019). Studying who responds to human capital development opportunities within the workplace may also be informative for the administration of public and social-sector training programs, which often report difficulty in attracting participation (Delfino et al., 2024).

Second, our results offer a proof of concept that the greatest beneficiaries of human capital development programs often fail to take advantage of the resources available to them.[7] When some types of workers are less likely to engage with workplace programs, recruiting the right people is likely complementary to programs that rely on self-selection (Oyer and Schaefer, 2011; Del Carpio and Guadalupe, 2022).

What is clear from our results is that treatment effect heterogeneity is significant. Failure to participate in programs by weaker employees likely contributes to the long-term, widespread productivity dispersion within firms that has been documented across many other settings (for an overview, see Hoffman and Stanton (2024); for healthcare, see Finkelstein et al. (2016), Currie and MacLeod (2017), Currie and MacLeod (2020), and Chan et al. (2022); for judges, see Coviello et al. (2014); for teachers, see Chetty et al. (2014); and for services, see Lazear et al. (2015, 2016)). Prior research has identified key drivers of differences in productivity, innovation, and compensation across workers and firms, such as management practices, managerial talent, capital formation, labor market concentration, and firm size (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Bloom, Lemos, Sadun, Scur, and Van Reenen, 2014; Larrain, 2015; Lazear, Shaw, and Stanton, 2015; Custódio, Ferreira, and Matos, 2019; Benmelech, Bergman, and Kim, 2022; Benson, Li, and Shue, 2019;

[7] Our survey also provides new context around the prevalence and characteristics of workplace programs, complementing studies of programs in particular contexts or industries (Rockoff, 2008; Jones et al., 2019; Chatterji et al., 2019; Reif et al., 2020).

Bandiera, Prat, Hansen, and Sadun, 2020; Friebel, Heinz, and Zubanov, 2022; Metcalfe, Sollaci, and Syverson, 2023; Benson, Li, and Shue, 2024). We show that variation in the administration of workplace programs may have profound consequences for the development and performance outcomes of workers in the lower tail of the productivity distribution.[8] This heterogeneity also speaks to the emerging evidence on the importance of managers for encouraging workers' career development and training (Hoffman and Tadelis, 2021; Minni, 2023; Diaz et al., 2024), in addition to the role of structured management practices in both attracting and retaining top-workers (Cornwell et al., 2021).

Third, our findings suggest that firms *must* consider the selection margin when testing workplace programs. While most work on selection underscores that firms can use design features to get advantageous selection or to exploit workers' behavioral biases (Larkin and Leider, 2012; Hoffman and Burks, 2020; Carter et al., 2019; Englmaier et al., 2021; Huffman et al., 2022), our findings suggest that selection effects are: (a) significant, and (b) difficult to predict ex-ante. This highlights the importance of running pilots with different recruitment and selection criteria ahead of broad deployment. Iconic recommendations on the econometrics of program evaluation suggest that randomization among program applicants is close to ideal for understanding potential outcomes (Heckman et al., 1997). Our results demonstrate that selection on who applies to a program can impact inference if the applicant composition changes when a program is deployed widely (List, 2022).[9]

[8]The early work in this area tended to be motivated by across-firm variation in management practices contributing to differences in TFP (Bloom and Van Reenen, 2007; Syverson, 2011; Gibbons and Henderson, 2012). Intra-firm experiments show that small changes in practices or incentives can lead to profound differences in output, which may contribute to across-firm variation (Friebel et al., 2017; Gosnell et al., 2020). These effects are likely even more pronounced when they interact with spillovers inside organizations (Mas and Moretti, 2009; Carrell et al., 2013; Bandiera et al., 2013; Herbst and Mas, 2015; Lazear et al., 2015; Cornelissen et al., 2017). Our findings contribute to an understanding of how variation in how firms do things contributes to performance heterogeneity (Englmaier et al., 2018; Bloom et al., 2019).

[9]Several earlier papers have examined endogenous entry/participation across different contexts (Karlan and Zinman, 2009; DellaVigna et al., 2012; Lazear et al., 2012), but many of these designs may be difficult to implement inside firms. Our design instead allows for simple variation in program recruitment procedures that enable tests of how treatment effects vary across selected samples.

2 Firm Setting

The study firm operates inbound sales call centers on behalf of several companies and brands, most of which are television, phone, and internet providers. Participants in the experiment are broadly representative of the 4 million U.S. workers in similar occupations. For example, average hourly earnings at the firm were about \$21 per hour in 2019, while customer service representatives, telemarketers, and miscellaneous sales representatives earned about \$23 per hour nationally and \$20 per hour in Utah, where the firm is located.[10]

The mentoring program occurred from January to December 2019. Our data tracks new hires' performance on the job after the conclusion of the program through early 2020. Sales agents answer incoming calls from potential customers and sell digital services with the goal of closing sales and up-selling premium service packages. Firm insiders report that learning the sales process (e.g., how to run credit checks for equipment lease compliance or determine whether callers qualify for regional sales promotions) and how to up-sell can be challenging for new hires.

When hired, sales agents begin a two-week training program, where they learn the sales process through lectures and by listening to other agents' live calls. Once agents complete their two-week training, they are allocated to a team and begin answering inbound sales inquires. Teams are typically comprised of 10–15 individuals, overseen by a (direct) sales manager, who is responsible for monitoring performance and troubleshooting issues faced by the agents. Individuals from the same hiring cohort can be allocated to different teams after training, however cohorts are recruited in service of selling a particular company's products. Agents eligible for the mentorship program were spread across seven different sales divisions, corresponding to different companies' brands or products.

This setting has several attractive features for studying the efficacy of mentorship. Most importantly, the firm provided us with individual-level performance measures for each sales

[10] These figures come from the 2019 5-year American Community Survey for SOC codes 43405, 41904, and 41309. To construct hourly earnings in the ACS data, we divide total individual income by the product of weeks worked last year and usual hours per week.

agent. Sales agents work independently on a call from start to finish, without subsequent hand-offs. Incoming calls are allocated to the next available agent within the appropriate division (each division receives calls from different phone numbers depending on the service being sold and the location of the callers, and the opportunities are then randomly allocated to agents in the division through the firm's call routing software). As such, agents do not have prior information about which calls may be more or less lucrative; i.e., they cannot sort into better opportunities. Agents generate revenue through each sale they make. The firm's focal productivity measure is revenue-per-call (RPC) because it allows managers to remove demand variation when comparing performance across workers. In addition, total revenue is important for workers, as the absolute amount of revenue generated impacts workers' commission pay. At the end of each week, the total amount of revenue generated is multiplied by an agent's commission rate. The commission rate is a coarse function of the agent's selling efficiency (determined by RPC and revenue per hour worked), relative to other agents in the same division. Commission rates range from 3% to 8%.[11] Multiplying the worker's revenue and commission rate determines their weekly commission pay. Sales agents also earn an hourly wage that begins above the federal minimum wage and increases with tenure.

3 Experimental Design

The experiment involves two high-level treatment conditions that were first assigned at the new-hire training class (cohort) level. Lower-level sub-treatments involving the assignment of mentors then occurred within each cohort. Training cohorts are specific to an office location and division. Cohorts joined the firm on a rolling basis during the experiment. We randomly assigned each cohort to either the Mandatory-Condition (probability 40%) or the Voluntary-Condition (probability 60%). Agents in the Voluntary-Condition were given the option to

[11] There is mild relative performance evaluation in this setting, and commissions increase at each quintile of selling efficiency. Helping another agent is unlikely to change relative rankings across quintiles, as the probability is small that any two agents are pivotal at the commission rate kink. Relative to settings with longer sales cycles (e.g. Oyer (1998) Larkin (2014)), incentives based on relative performance reset weekly.

opt in or out of mentoring. Those who opted out did not receive a mentor. Agents in the Mandatory-Condition and those in the Voluntary-Condition who opted in were randomly assigned a mentor, or not, according to the following rule: if the supply of available mentors was greater than 50% of the cohort size, then approximately half of the agents would be assigned a mentor (the firm requested that we randomly allocate mentors to more agents when possible, e.g., rounding up for an odd number of agents in a cohort); otherwise, the available mentors would be assigned at random to those eligible to receive a mentor.[12] The pairing of mentors and new hires always occurred at random.

Figure 1 displays the allocation of cohorts and agents to the different conditions and treatments in the experiment. There were 591 program-eligible sales agents spread across 52 new hire cohorts.[13] Twenty-one cohorts and their 264 sales agents were allocated to the Mandatory-Condition, whereas the other 31 cohorts and 327 sales agents were allocated to the Voluntary-Condition. Among the agents in the Mandatory-Condition, 127 agents (48%) were randomized to receive a mentor, and the remaining 137 were not. In the Voluntary-Condition, 272 agents (83%) chose to opt in, of which 155 agents (57%) were randomized to receive a mentor, and the remaining 117 were not. The remaining 55 agents (17%) in the Voluntary-Condition chose to opt out of receiving a mentor.[14]

3.1 Timeline for Administering the Program and Communicating Treatment Allocations

Prior to starting the two-week training protocol, each cohort was allocated to either the Mandatory- or the Voluntary-Condition, and the staff administering the program was made aware of the cohort's assignment. All new hires were asked to complete a survey on the first day of training, which asked about their personality traits, work styles, and work experiences (specifically, whether they had call center and/or sales work experience). We use these survey

[12]When mentor supply fell below 50% of the number of eligible new hires, the most common reason was conflicting obligations to mentor other cohorts in the same division or office.

[13]Our prior working paper version reports 53 cohorts assigned to treatments. We erroneously coded one cohort that had no available mentors as eligible for the experiment.

[14]Across months, opt-out rates range from 6% to 43%, with no obvious time trend.

responses to identify the characteristics of individuals who opted into versus opted out of mentoring.

For cohorts in the Mandatory-Condition, agents were either randomly assigned a mentor or not based on the assignment rule described above. For cohorts in the Voluntary-Condition, the staff described the mentoring program to the newly hired agents and told them they could either opt in or opt out of participating. The agents were told that a randomly selected subset of those who opted in would receive a mentor at the end of the training period. The staff explained that the supply of mentors was limited and an outside research team would help with the randomization to ensure fairness in the assignment.[15] To avoid peer influence in program participation (Dahl et al., 2014), agents were asked to write on a piece of paper whether they wanted to opt in or out of the mentoring program, making their decision anonymous to their peers. Among those who opted in, agents were either randomly assigned a mentor or not, based on the assignment rule described above. Agents assigned a mentor were informed of this assignment by the within-firm staff during the last days of their training. To reduce the possibility of discouragement among agents in the Mandatory-Condition who were not assigned a mentor, the staff did not initially inform them about the mentorship program. If agents inquired about why they were or were not assigned a mentor, the staff explained that the mentor supply was limited and that available mentors were randomly allocated to new hires.[16]

Across all treatment conditions, the two weeks of training remained exactly the same for all agents, regardless of their treatment assignment. After the two weeks of training, new hires graduated to work as regular agents, began taking customer calls, and had measurable sales productivity metrics. It was only then that meetings with mentors commenced.

[15] The staff members were asked to read the following statement to new hires in the Voluntary-Condition: "We have recently begun a mentorship program to help newly hired sales agents when they begin working on the sales floor. Agents who opt into the program and are chosen by [the research team] will be assigned a mentor. Your mentor will approach you during your first week on the sales floor to initiate the mentoring relationship. The program will run from your first week on the sales floor to your fourth week on the sales floor, and you and your mentor will meet once a week to discuss your progress."

[16] The staff reported to the authors on multiple check-in calls that they found no evidence of discouragement among the agents who did not receive a mentor.

To facilitate meeting coordination, the firm built specific times to meet into mentors' and protégés' schedules. The mentoring relationships lasted for four weeks in most cohorts (the study's pilot program used a six-week design, which we discuss in Section 3.4).

Mentors and protégés met once per week for approximately 30 minutes and completed a worksheet. They were free to discuss any topic, but the worksheet had to be completed for the mentor to receive credit for the meeting (as described below). Records of meeting occurrences and completed worksheets were kept by the staff and given to us. Shortly after their final week of meetings, protégés were asked to complete a post-mentorship survey about their experience. Although completion rates for the final survey were low, we use the data to provide insight into whether meetings continued after the formal program and whether agents viewed the experience as beneficial.

3.2 Identifying Mentors

The firm's staff sourced mentors by announcing to incumbent sales agents that a mentoring program for new hires would occur and that agents could volunteer to be a mentor. The staff directly asked some promising candidates to participate. Agents who the staff felt were not suitable to be mentors were excluded. Mentors were given two main incentives to participate. First, in exchange for each pre-scheduled, confirmed meeting they held with their protégé, they received internal currency ("kudos" dollars) worth approximately $10. Second, incumbent sales agents were told that effective mentoring would help demonstrate leadership potential for future promotion considerations. It is important to note that mentors in this setting had no formal supervisory role; they were more experienced peers who had proven track-records of sales success.

Mentors were always randomly assigned to protégés. Table I.A.1 shows that the observable characteristics of the mentors—age, gender, marital status, and tenure—are similar across the Voluntary- and Mandatory-Conditions, meaning that endogenous matching of mentors to protégés or homophily do not explain differences in performance across the two high-level treatment conditions. Mentors were not informed about which condition their protégés were in.[17]

3.3 Hold-Out Cohorts to Test for SUTVA Violations

There were 217 agents hired throughout the experiment in cohorts that were ineligible for the mentorship program. Ineligibility largely arose because these cohorts entered at times when mentor supply was lacking. Insufficient mentor supply typically occurred when the firm hired a new cohort soon after another one finished training, but in some cases projected call volumes relative to available staffing meant that potential mentors would not have time to meet with new hires. Agents in these cohorts form hold-out groups that were not informed about the mentorship program. Variation in treatment eligibility at the cohort level allows us to test for discouragement effects in the control group and other possible violations of the Stable Unit Treatment Value Assumption (SUTVA). Although these hold-out cohorts were not randomly assigned, they have similar characteristics as program-eligible cohorts in the same division and office. We leverage these hold-out cohorts to compare the productivity of hold-out new hires to the productivity of non-treated agents in program-eligible cohorts, showing that SUTVA violations were unlikely (see the Internet Appendix, Section I.A).

3.4 Pilot Data

We piloted our design in the firm from January to May of 2019 to ensure we could logistically implement the program. The pilot surfaced several virtues of the program while assuaging feasibility concerns: (i) there was sufficient interest among seasoned agents to mentor new hires, (ii) the firm could schedule meetings between mentors and protégés, (iii) mentors

[17] Mentors were not designated exclusively to either the Mandatory- or Voluntary-Condition, so their first protégé could have been in one condition and their second protégé could have been in the other condition. Mentors generally only mentored a single protégé at a time, but there were instances where a mentor was assigned to multiple protégés at once. This only occurred when mentors were in short supply and the firm's internal staff felt that the mentors could effectively handle the assignment. In all cases, though, to facilitate meeting coordination, the firm built specific times to meet into mentors' and protégés' schedules, and mentor-protégé pairs always met individually, meaning the protocol was the exact same from the point of view of the protégé.

and protégés would engage with the protocol as designed, (iv) anecdotal evidence indicated that protégés felt they benefited from the mentorship, and (v) there were no indications of discouragement among non-mentored agents.

As a result, we moved forward with the experimental design described thus far, which varied from the design of the pilot in only two ways. First, to accommodate scheduling, we changed the duration of the mentorship program from five meetings over six weeks (with a gap in week five) to four meetings over four weeks. Second, at the beginning of the pilot, the allocation of cohorts to the Mandatory-Condition and Voluntary-Condition was determined by the location of each cohort; i.e., all cohorts at one office were allocated to one condition, and those at the second office were allocated to the other. This allocation was chosen to limit potential spillovers between the Mandatory- and Voluntary-Conditions (e.g., workers potentially talking about the choice to opt in). Within each condition, the firm's staff observed no discussion of program logistics among new hires or spillover effects within or across cohorts. There were also no complaints from agents in the Voluntary-Condition who requested but did not receive a mentor. Accordingly, we determined the risk of spillovers *across* conditions was small and the logistics were feasible such that we could randomize Mandatory- and Voluntary-Condition assignment within offices as well.[18]

No other changes were made between the pilot period and the later cohorts. The pre-registration text was finalized after the pilot and is documented in the Internet Appendix (see Section I.B).[19]

Based on power calculations and the hiring projections given to us by the firm, we

[18]In our previous experience conducting experiments within this setting, we found no evidence of spillovers from one treatment group to another among sales agents within the same office (Sandvik et al., 2020). In particular, we leveraged data from sale agents in a separate office that was not part of (or informed of) the experiment, and we found that their trends in sales performance mirrored those of the agents in the control group who were aware of the experiment (those located in the two participating offices), but who were not treated with any stimulus to alter their behavior. In addition, in that setting we found significant differences in treatment effects between the conditions that nudged agents to share best practices, and those that did not—even though the printed prompts to share best practices (a physical worksheet with questions) could have easily been disseminated across treatment groups. In the Internet Appendix, we present tests for contamination and spillovers outside of the experimental treatments.

[19]Instructions given to mentors and the mentor-protégé worksheet can be found in the Internet Appendix (see Section I.C).

expected the firm to hire 619 agents across 46 cohorts after the pilot period (May to December of 2019). The actual hiring at the firm was much less frequent and intense, with the firm only bringing on 276 agents across 27 cohorts that were eligible for the mentorship program. We were not able to extend the mentorship program into 2020, as COVID-19 forced all employees to work remotely. Since the firm's actual hiring behavior was substantially less intense than expected, and given the similarity between the experimental design in the pilot period and the pre-registered period, our empirical analyses include the 315 agents and 25 cohorts from the pilot to improve statistical power. We detect no differences in treatment effects or imbalance in worker characteristics between the pilot cohorts and those from the post-pilot period (see Table I.A.2).

3.5 Balance Across Treatments

Agent characteristics are balanced across the conditions of the experiment and across the treatment statuses within each condition for those agents eligible for randomization. Table 1 Panel A displays cohort-level balance tests for the Mandatory-Condition compared to the Voluntary-Condition (the top level of randomization). There are no significant between-condition differences in average agent age, gender, marital status, hiring score (recruiters' evaluation of the worker's suitability for the position), and referral status. The average agent age in both groups is about 23 years old, women make up 43% of the agents in the Mandatory-Condition and 40% of agents in the Voluntary-Condition, and 13%–16% of agents are married in the two groups. The average hiring scores (which have a maximum value of 1) are 0.83 and 0.85, respectively. These scores are based on the recruiters' perceptions of applicants' sales experience, ability to adhere to the sales process, self-awareness, competitiveness, and personal motivation. We also report adjusted hiring scores, which take into account some recruiters' relative scoring leniency compared to others—akin to curving grades received from one professor versus another. Throughout our analysis, we use the adjusted hiring score because it is a better predictor of opting out of the program relative to the raw hiring scores, but our results are not sensitive to the use of raw hiring scores, which we discuss in Section 4.3.1.[20]

Panel B of Table 1 considers the second level of randomization, the allocation of mentors to new hires *within* the Mandatory-Condition or Voluntary-Condition. Columns (1) and (2) show the agent-level average characteristics in the Mandatory-Condition for those who did and did not receive a mentor, respectively. These two groups are similar in age, gender, marital status, hiring scores, adjusted hiring scores, and referral status. Columns (3) and (4), and the associated p-values show that agents assigned mentors and those that were not in the Voluntary-Condition, conditional on opting into the program, are similar across these observable characteristics as well.[21] We defer discussion of differences between agents who opt into and out of the program to Section 4.3.1.

4 Estimation and Results

4.1 Treatment Effects on Productivity and Selection Into Mentoring

We estimate differences in productivity by high-level treatment condition (Mandatory or Voluntary) and low-level sub-treatment cell (assigned a mentor, not assigned a mentor, or opted out). We refer to agents assigned a mentor as "mentored," which we use to denote treatment assignment in an intention-to-treat framework. Our main productivity outcomes of interest, $y_{i,t}$, are total daily sales revenue (Revenue) and daily revenue-per-call (RPC). Total daily revenue directly relates to the firm's profitability, while accounting for the opportunity cost of time spent meeting a mentor. RPC captures selling efficiency on a given

[20] There are 15 recruiters in the data. Some recruiters systematically give higher scores than others conditional on the performance of the workers they evaluate. We find this relationship for workers who are not part of the experiment, and we account for it using a procedure that adjusts for the stringency or leniency of each recruiter. Using data on workers outside of the experiment, we recover recruiter relative leniency by regressing raw hiring scores on productivity (specifically, the inverse hyperbolic sine of revenue and the inverse hyperbolic sine of revenue-per-call), recruiter fixed effects, brand fixed effects, and time fixed effects. We then shrink the recruiter fixed effects (that are net of the productivity adjustment) using the procedure in Lazear et al. (2015). We subtract the adjusted recruiter fixed effects from the raw hiring scores of workers in the experiment to return the adjusted hiring scores.

[21] We also check for balance across assignment to divisions based on estimated division-level productivity for workers outside the experiment. Table I.A.3 shows that assignment is balanced on the productivity metrics of non-mentor-eligible new hires (from hold-out cohorts) across divisions.

opportunity. We also form a composite index of productivity measures that incorporates two additional—albeit less central—performance measures that are tracked by the firm. These are adherence, which captures on a zero to one scale how closely agents adhere to their preset schedules (i.e., are available to take calls when they are supposed to be on the phones), and revenue-per-hour, which scales total revenue by hours worked. We pre-registered a natural specification to capture percentage changes in revenue and RPC within a cohort. Cohort fixed effects sweep out division-level differences in baseline revenue and RPC (because cohorts are assigned to a single division). To account for the fact that some agents have days with zero sales revenue, we use the inverse hyperbolic sine transformation (IHS). Parameter estimates can be interpreted as approximate percentage changes.[22] We use a sample of agent-day productivity data for all program-eligible agents in their first two months on the job after completing training (estimates for months three through six are discussed in Section 4.1.4).

Our first specification comes from a linear regression, fit separately for the Mandatory- and Voluntary-Conditions, on the sample of agents who were eligible to be assigned a mentor (e.g., they did not opt out of the program):

$$\mathrm{y}_{i,t} = \alpha + \beta_1 \mathrm{Mentored}_i + \gamma_j + \varepsilon_{i,t}. \tag{1}$$

The variable $Mentored_i$ is an indicator taking the value of one for agents who were randomly assigned to receive a mentor. The t subscript denotes the calendar date, and γ_j is a cohort fixed effect at the unit of randomization that absorbs product- and brand-level differences. We cluster standard errors by cohort for those workers entering the experiment after the pilot-period and at the pilot-period-by-office level for those workers entering during the pilot (recall that the pilot program entailed assignment of the Mandatory- and Voluntary-

[22]The inverse hyperbolic sine transformation was not under consideration at the time we pre-registered using dependent variables in logarithms, as at the time we were unaware of the fact that workers occasionally experienced zero-revenue days. Our results are similar if we use the natural logarithm of one plus revenue or one plus RPC. While the results are qualitatively unchanged, we nonetheless provide the results using the logarithmic transformation in Table I.A.4.

Conditions at the office level).

4.1.1 Treatment Effects in the Mandatory-Condition

Due to random assignment of mentors to some Mandatory-Condition agents and not to others, when estimating Equation (1), the parameter β_1 is the average treatment effect of receiving a mentor across the *entire* population, not the treatment effect conditional on opting into the program. We tabulate the results in Columns (1) and (2) of Table 2 for IHS(Revenue) and IHS(RPC), respectively. In both columns, we estimate positive effects that are statistically significantly different from zero. The estimate in Column (1) implies that mentored agents in the Mandatory-Condition generated 18.6% ($= e^{0.171} - 1$, p-value $=$ 0.002) more daily sales revenue than their non-mentored peers. The estimate in Column (2) implies that mentored agents generated 11.9% ($= e^{0.112} - 1$, p-value $= 0.003$) more in revenue-per-call.[23] To understand the magnitude of the estimates, we compare them to the baseline gap between new hires and experienced agents. In our setting, the average experienced worker is about 30% more productive than the average new hire (see Table I.A.5 in the Internet Appendix), suggesting that the program accelerates on-the-job learning, albeit it fails to fully close the average productivity gap between new hires and experienced agents.[24]

4.1.2 Treatment Effects in the Voluntary-Condition

Next we estimate Equation (1) on the sample of Voluntary-Condition agents who opted into the program. If the employees who are likely to benefit the most from mentorship are also those who opt into the program, then we would expect to see treatment effects in the Voluntary-Condition that exceed those in the Mandatory-Condition. If, however, the employees who are likely to benefit the most are also those who opt out of the program, then

[23] Differences between revenue and RPC estimates arise from differences in hours and/or calls. Mentored agents in the Mandatory-Condition, on average, field 0.4 more calls per day (p-value $= 0.134$) and they work 0.2 more hours per day (p-value < 0.01) than their non-mentored peers.

[24] Our effects are smaller than those documented from the introduction of technology, as Brynjolfsson et al. (2023) find that access to a generative A.I.-based tool increased the productivity of newly hired call center agents doing customer service work by 34%. However, our effect sizes are roughly equal in magnitude to the 13% lift associated with remote work found in Bloom et al. (2015).

treatments effects in the Voluntary-Condition will be smaller than those in the Mandatory-Condition. Due to random assignment of mentors to some opt-in agents and not to others, the parameter β_1 in the Voluntary-Condition is the average treatment effect of receiving a mentor conditional on selection into program participation. The results are in Columns (3) and (4) of Table 2 for IHS(Revenue) and IHS(RPC), respectively. In both columns, the estimated effects are statistically indistinguishable from zero. Random assignment to be mentored had a negligible effect on the productivity of workers in the Voluntary-Condition conditional on their opting into the program. These estimates are much smaller than those in the Mandatory-Condition. Had the analysis been conducted only among those who selected into randomization, which is typical for many RCTs across disciplines ranging from medicine to economics, we would have falsely concluded that the program was not effective in the population. Instead, these results suggest that different procedures for administering the program can change inference. We turn now to assessing why estimates differ across conditions.

4.1.3 Self-Selection in the Voluntary-Condition

How much of the difference in treatment effects between the Mandatory- and Voluntary-Conditions arises from selection into participation? To provide color on just how much selection bias may be present, we estimate how workers' baseline productivity varies as a function of whether they volunteer to participate. Specifically, we compare non-mentored agents in the Voluntary-Condition who opt into the program with those who opt out using the following regression:

$$\mathrm{y}_{i,t} = \alpha + \beta_1 \text{Voluntary Opt-Out}_i + \gamma_j + \varepsilon_{i,t}. \tag{2}$$

The variable *Voluntary Opt-Out*$_i$ is an indicator for agents who opted out of the mentorship program when given the opportunity. The parameter β_1 captures the difference in productivity between agents who did, and those who did not, opt out of the program. Results for revenue and revenue-per-call are reported in Columns (5) and (6) of Table 2, respectively. The estimates imply that opt-out agents generated 30.9% ($= e^{-0.369} - 1$, p-value = 0.003) *less* revenue per day than non-mentored, opt-in agents and had 23.2% ($= e^{-0.264} - 1$, p-value = 0.001) lower productivity on a per-call basis. The agents who opted into program participation were significantly more productive, on average, than those who opted out.

4.1.4 Pooled Estimates, Additional Productivity Measures, Multiple Tests, and Long-Term Outcomes

In Columns (7)–(9) of Table 2, we estimate all three effects of interest simultaneously in a single model that includes all mentor-eligible agents across both the Mandatory-Condition and the Voluntary-Condition. The model is:

$$
\begin{aligned}
\mathrm{y}_{i,t} = \alpha + \beta_1 \text{Mentored}_i + \beta_2 \text{Mentored}_i \times \text{Voluntary}_i \\
+ \beta_3 \text{Voluntary Opt-Out}_i + \gamma_j + \varepsilon_{i,t},
\end{aligned}
\tag{3}
$$

where *Mentored*$_i$ indicates the agent was randomly assigned to receive a mentor, *Voluntary*$_i$ equals one for agents in the Voluntary-Condition, and *Voluntary Opt-Out*$_i$ equals one for agents in the Voluntary-Condition who opted out of the mentorship program. In this model, β_1 captures the treatment effect of mentorship among agents in the Mandatory-Condition, β_2 captures the difference in treatment effects among opt-in agents in the Voluntary-Condition, relative to mentored agents in the Mandatory-Condition, and β_3 captures the selection effect among non-mentored agents in the Voluntary-Condition. The baseline effects for the Mandatory- and Voluntary-Conditions are absorbed by the cohort fixed effects, which also control for differences in productivity that are specific to the time when agents entered the firm and the differing products sold.[25] Pooling the models allows us to test whether treat-

[25] All of our pre-registered specifications include cohort fixed effects, as we expected that between cohort variation would significantly increase minimum detectable effect sizes. With cohort fixed effects, calendar time and elapsed time since hire are co-linear. In a balanced panel with a short time window, cohort-fixed effects absorb time-fixed effects. We show in Figure I.A.1 that our results are robust to the inclusion of date fixed effects as well as to the use of several other alternative specifications.

ment effects differ between conditions.

The pooled model results for IHS(Revenue) and IHS(RPC) are reported in Columns (7) and (8), respectively. In the top row, the productivity treatment effects for the Mandatory-Condition are identical to the prior estimates. In the second row, the point estimates of productivity differences for those who opt out in the Voluntary-Condition are similar to the prior estimates, but they are not identical because the sample changes relative to the columns that focus only on comparing unmentored agents. The third row shows that the treatment effect of receiving a mentor in the Voluntary-Condition is statistically different than the treatment effect of receiving a mentor in the Mandatory-Condition. The bottom row of Table 2 reports tests of the null that treatment effects are zero for those mentored in the voluntary program, as their treatment effects are the sum of the coefficients on *Mentored* and *Mentored* $\times$ *Voluntary*.

To capture the suite of productivity measures, in Column (9) we repeat the pooled analysis with an alternative dependent variable that factors in adherence to schedule and revenue-per-hour as additional outcomes. We construct a standardized, weighted summary *Index* of all performance metrics (see Anderson (2008)): IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence. The measure is normalized to have mean zero and unit standard deviation for non-mentored agents in the Mandatory-Condition.[26] We continue to find that the program generally raised productivity when it was mandatory, that the program had no effect in the Voluntary-Condition, and that Voluntary-Condition participants who opted in were stronger than those who opted out. The economic magnitudes of the point estimates in Column (9) must be interpreted differently from the other columns, as they are in standard deviation units relative to the control mean. Thus, treatment in the Mandatory-Condition *raised* over-

[26]The summary index approach has been used to evaluate education interventions when there are multiple potential outcomes (Deming et al., 2014). The procedure first demeans and standardizes each individual outcome by the control group standard deviation (in this case, non-mentored agents in the Mandatory-Condition). The index is then the weighted sum across inputs, where the weights come from the inverse of the covariance matrix of the standardized measures, akin to the approach in generalized least squares. Anderson (2008) argues that this approach has three advantages: (i) it allows for a single test rather than multiple tests across different outcomes, (ii) it is a test of whether a program has a general effect, and (iii) the tests are potentially more powerful than multiple tests with marginal significance.

all productivity by 0.145 standard deviations, and agents in the Voluntary-Condition who opted out have productivity that is 0.141 standard deviations *lower* than those who opted in.

We also correct for multiple hypothesis testing using a second suggestion by Anderson (2008), where we report sharpened q-values that are analogous to a p-value after adjusting for the False Discovery Rate (FDR). The q-values, reported in Table 2 in brackets below the standard errors, indicate that inference regarding our main point estimates is robust to holding fixed the proportion of false positives as the number of tests increases.[27]

In Table I.A.6, we show that about 45% of the point estimates from months 1–2 persist through months 3–6 for mentored workers in the Mandatory-Condition, while the effect of having a mentor in the Voluntary-Condition remains close to zero. The longer-term point estimates have larger standard errors relative to the effects at 1–2 months of tenure for two primary reasons: a) there is an increase in residual variation as agents gain experience, causing productivity to fan out, and b) there are fewer agents who remain at the firm over longer time horizons. While we lose precision, the pattern of estimates suggests that the mentorship program helped treated workers in the Mandatory-Condition over the longer-term.

4.2 The Mentorship Program Did Not Impact Worker Retention

Call centers have notoriously high levels of attrition (Hoffman et al., 2017), and retention is a key performance metric for the HR executives at the firm. To estimate retention effects from the mentorship program, we use data with a single observation per unique mentor-eligible agent among those who completed training,[28] and we create an indicator variable $Tenure_{30}$ ($Tenure_{60}$) that equals one for agents who remain with the firm for at least thirty (sixty) days

[27]The q-values are adjusted for tests on all regressors reported in Columns (7)–(9) of Table 2, as well as all regressors reported in Columns (7)–(9) of Table I.A.6, which capture the long-term treatment and selection effects of mentorship in months 3–6 of agents' tenure with the firm. The estimated sharpened q-values are conservative in our case because they do not account for the positive correlations across tests.

[28]The results are similar if we include individuals who did not complete training into the retention estimations.

after their hire date, and zero otherwise. We then re-estimate each of the models specified by Equations (1)–(3) with these two tenure achievement indicators as the dependent variables.

In Table 3, we find no evidence that mentorship impacted agents' retention, although agents who opt out of the program in the Voluntary-Condition are less likely to achieve sixty days of tenure than are non-mentored agents who opt in (albeit the estimate is noisy). There are no discernible retention effects among agents who were mentored relative to those who were not at these horizons or at longer horizons (see Table I.A.7 in the Internet Appendix).[29]

To further analyze the relation between mentorship and retention, we plot the distribution of completed tenure for mentored agents in Figure I.A.2. Specifically, we plot the distribution of completed tenure, in years, for each mentored agent in the Mandatory-Condition (solid line) and for each mentored agent in the Voluntary-Condition (dashed line).[30] Comparing the distributions we see that mentored agents in the Mandatory-Condition realize slightly higher levels of retention, relative to mentored agents in the Voluntary-Condition. Formal tests of mean, standard deviation, and distribution differences do not reject the null that the two groups realize the same tenure outcomes.[31] As the distribution of completed tenure among mentored agents does not differ between the Mandatory-Condition and the Voluntary-Condition, it is unlikely that retention differences drive our main productivity findings. Given that mentorship does not appear to impact agents' retention, it is unlikely that the differences in productivity treatment effects between conditions are driven by differences in attrition, a point we return to in Section 4.4.2.

[29]One possible, albeit speculative, explanation for the lack of a retention effect among agents in the Mandatory-Condition—despite their improved productivity—is that exposure to a top-performer (i.e., their mentor) may have increased their awareness of how much they had yet to learn. This could have discouraged them about their long-term career prospects at the firm, even if their performance was accelerating faster than that of their non-mentored peers.

[30]For all agents, the completed tenure is calculated as the difference between their hire date and the date of the last day they are observed in the data, divided by 365.

[31]If we conduct a t-test on completed tenure between the two groups, the averages are 0.391 and 0.394 with a p-value of 0.9581. Similarly, we cannot reject the null that the standard deviations of these two distributions differ (p-value $= 0.2446$), nor can we reject the null that the distributions differ via the rank-sum test (p-value $= 0.2047$).

4.3 Selection and Treatment Effect Heterogeneity

We study the mechanisms underlying the differing productivity treatment effects between the Mandatory-Condition and the Voluntary-Condition throughout the rest of the paper. While our initial evidence suggests that heterogeneous treatment effects and self-selection might be the cause of the differences, these channels are likely not big enough to explain the totality of productivity treatment effect differences between the two conditions. We also explore several alternative channels that could have caused the treatment effects to differ. Because our experiment was done in the field, the greater effectiveness of mentorship in the Mandatory-Condition could have been caused by logistical differences, framing issues, or other unobserved factors—all of which could be expected to (similarly) surface in any real-world intervention or test involving the implementation of a mandatory program versus a voluntary program. We will use differences in outcomes for agents *within* the Mandatory-Condition to show the presence of treatment effect heterogeneity that varies with selection probabilities, and our evidence on other channels is suggestive of a program framing effect.

If differences in treatment effects across conditions are driven by self-selection, then there must be treatment effect heterogeneity such that those who are most likely to opt out of the program have the largest gains when they receive mentorship. We now turn to understanding selection before evaluating the other explanations.

4.3.1 Differences Between Agents who Opt Out and Those who Opt In

We first consider how agents who opt out differ from those who opt into the program. We restrict the sample to the 365 agents in the Voluntary-Condition who were given the choice to opt in or out, and we estimate logistic regressions of a *Voluntary Opt-Out* indicator on worker characteristics.[32]

The main conclusion from this analysis is that low hiring scores, assigned during the

[32] In this analysis, we include agents who did not complete training, in which case they do not have productivity data. This accounts for the difference in unique worker counts between this sample and that reported in Figure 1. Results are similar when we use only workers who completed training to examine the determinants of opting out.

interview stage, are the best predictors of opting out of the program. Worker demographics from the firm's personnel records and personality traits (obtained via on-boarding surveys) do little to explain program participation. Failure to complete the on-boarding surveys also predicts opting out of the program, but this is not a characteristic that is measured up-front. These results can be seen in Table 4, Columns (1)–(4), which report marginal effects from logit models predicting the opt-out decision. In Column (1), we report no difference between agents that opt in and opt out based on age and marital status. The impact of gender is significant at the 5% level in Column (1), but it loses explanatory power and decreases substantially in magnitude when controlling for additional covariates. Participation decisions do not depend on an agent's location (a fixed effect for one office compared to the other) or whether an existing employee referred the agent (following Friebel et al. (2023)). Participation decisions also do not depend on whether the agent had prior sales experience (which we collected from the new hire survey for 341 agents) or fixed effects for the agents' assigned division. Personality characteristics, also collected from the new hire survey, are weak predictors of the opt-out decision. However, we find that agents who did not complete the new hire survey and those with prior call center experience have a higher propensity to opt out, as reported in Column (4).

The best predictor of the opt-out decision is contained in recruiters' assessments of the new hires' suitability for the job. We find that agents with higher adjusted hiring scores—interview scores net of recruiter leniency (see Section 3.5)—are more likely to opt into the program. As an agent's hiring score is given to them by the recruiter who interviewed them for the job, this suggests that recruiters' assessments of agents' suitability for the job can predict their program engagement.[33] Computing marginal effects from the logit model, which are reported in the table for a unit change in each regressor, we find that an increase in the adjusted hiring score of 0.10 (approximately the interquartile range in the sample) is

[33] We are missing hiring score data for 25 agents, so we set their hiring scores to zero and include an indicator variable that they had missing data. We find similar results in Table I.A.8 when using raw (non-adjusted) hiring scores.

associated with a 6.9 percentage point decrease in the likelihood that the agent opts out of the program.

We also assess the extent to which the predictors of program opt-out explain variation in agent productivity. Using a sample of agents from the Voluntary-Condition who were not mentored, we regress realized productivity on the factors that potentially explain program participation. Column (5) displays the baseline productivity regression results controlling only for agent demographics, hiring scores, referral status, and cohort fixed effects (which absorb the location dummy). The coefficient on *Adjusted Hiring Score* is positive and statistically significant. A one standard deviation change in the adjusted hiring score (approximately 0.07 units), yields a 19% change in revenue. This suggests that both the opt-out decision and the observable agent characteristics (that predict opting out) help to explain on-the-job productivity. Column (6) adds data from the new hire survey and the *Missing Survey* dummy. The coefficient on *Adjusted Hiring Score* is larger in magnitude—even when other characteristics are included. The results are similar in Column (7) when the dependent variable is IHS(RPC), showing that hiring scores predict on-the-job performance. At this firm, and likely in others, workers with low pre-hire assessments are less productive than other workers with more favorable evaluations. As we show later, human capital development programs can help remediate this lower level of initial productivity.[34]

4.3.2 Heterogeneous Treatment Effects For Agents Who Are Likely to Opt Out of the Program

Next, we conduct tests of heterogeneous treatment effects within the Mandatory-Condition. These tests are robust to framing or logistical differences that may drive other variation between the Mandatory- and Voluntary-Conditions. Here we ask whether the largest *individual* treatment effects accrued to mentored workers in the Mandatory-Condition with the greatest

[34] Readers may wonder why the firm would hire applicants with low interview scores. The seasonal nature of subscription sales requires immediate workforce capacity, and the firm often would need to take a set of applicants as given and pick the best among them rather than continue to recruit better agents to fill out a training class.

likelihood of opting out of the program. To do this, we use the coefficient estimates in Column (1) of Table 4 and the characteristics of workers in the Mandatory-Condition to impute opt-out propensity scores for those workers. We classify agents in the Mandatory-Condition as either $High_{Opt}$, if their opt-out propensity score is in the top tercile of the distribution, or Low_{Opt}, if their opt-out propensity score is in the bottom two terciles of the distribution. We use terciles rather than the 17% opt-out rate in the Voluntary-Condition because (i) the individual propensity scores are less than 1, implying we need more workers to yield the total number of those who opt out in the Voluntary-Condition and (ii) we run into power issues due to small samples if we use fewer agents or a finer partition. If treatment effects are monotonic in the propensity to opt out of treatment, then these choices are conservative. We then estimate Equation (1) on these subsets of the data alongside pooled models that allow us to test for differences in treatment effects between workers with high and low propensities to opt out of participation.

We find that agents in the Mandatory-Condition with a high estimated likelihood of opting out had a significantly greater treatment effect of mentorship than did their peers who were less likely to opt out, as reported in Table 5. The estimate in Column (1) shows that agents who were most likely to opt out of the program had revenue gains of over 38% ($= e^{0.324} - 1$), whereas the estimate in Column (2) shows that other agents had estimated gains of about 7%. The pooled estimate on $Mentored \times High_{Opt}$ in Column (3) rejects equality of the treatment gains within the Mandatory-Condition between high and low opt-out agents, providing evidence in favor of heterogeneous treatment effects. Column (3) also shows that agents in the highest tercile of the opt-out propensity distribution are about 21% ($= e^{-0.239} - 1$) less productive than other agents in the Mandatory-Condition. The results in Columns (4)–(6) report similar patterns when using IHS(RPC) as the dependent variable.[35]

[35] We also pre-registered a procedure for estimating heterogeneous treatment effects that yields larger estimates for those who opt out. We discuss this procedure in the Internet Appendix (see Section I.D). The results of the pre-registered estimations provide additional evidence that the treatment effects of mentorship are greatest among agents who are most likely to have opted out of the program. Our pre-registered estimates of treatment gains for opt-out agents are larger than those given here because that estimator imposes that the treatment effects for agents who opt in are constant across the mandatory and voluntary programs.

We note that the high opt-out propensity agents here are again less productive than those who are more likely to participate in the program, consistent with self-selection.

The estimates in Table 5 imply that the inclusion of workers who would have likely opted out of the voluntary program raised aggregate treatment gains in the Mandatory-Condition by 6% for revenue and by 5% for revenue-per-call. These figures come from taking the treatment effects of the program for top-tercile agents (38% revenue gains and 30% gains in RPC) and multiplying them by the actual opt-out rate in the Voluntary-Condition, 17%. If we assume that the actual treatment effect in the Voluntary-Condition among opt-in agents is zero, adding opt-out workers to those eligible for treatment would close approximately 34% (43%) of the gap in treatment effects between the Mandatory- and Voluntary-Conditions for revenue (RPC).[36]

To summarize, we find evidence of both self-selection—as higher productivity workers are more likely to opt into the mentorship program—and heterogeneous treatment effects—as the mentorship program helps some agents more than others. These channels explain part, but not all, of the differences in treatment effects between the mandatory and voluntary programs. We discuss in Section 4.4 that framing effects provide the most likely explanation for the remaining differences between conditions.

4.4 Framing Effects and Other Explanations for the Remaining Differences in Treatment Effects

Here we consider several potential alternative explanations for the remaining differences in treatment effects between the mandatory and voluntary programs. Framing effects appear to be the most likely of the candidates we consider.

The approach in Table 5 allows the treatment effects to differ across the treatment conditions for agents who are likely to opt in, and we find that the treatment effects for these agents are modestly positive in the Mandatory-Condition. The pre-registered approach is sensitive to this variation among agents who are likely to opt in, so we prioritize the estimator based on the propensity score that does not impose this restriction.

[36]For revenue, (38% × 17%) / 19% = 0.340. For RPC, (30% × 17%) / 12% = 0.425. We note that these are conservative estimates, as the propensity scores are measured with error, so we use estimates that average effects over the top third of the distribution of scores, covering more workers than the actual opt-out rate.

4.4.1 Framing Effects

It is possible that the mandatory framing of the program in the Mandatory-Condition caused agents to infer something about its value and "buy in" or engage, while agents in the Voluntary-Condition may have perceived the program as optional, reducing buy-in and engagement. To test for differences in compliance or buy-in between the Mandatory- and Voluntary-Conditions, we tabulate meeting completion rates between mentor-protégé pairs in Table 6. Of the 127 agents assigned to mentorship in the Mandatory-Condition, 18 never completed a recorded meeting with a mentor, while 25 of the 155 treated agents in the Voluntary-Condition never met with their mentor. Mandatory-Condition protégés completed both more of their scheduled meetings (2.31 versus 2.11) and had a higher meeting completion ratio (74% versus 64%).[37] The values in Column (4) show that meeting completion rates are even larger among agents in the Mandatory-Condition who likely would have opted into the voluntary program (those with low opt-out propensity scores), suggesting that similar agents have different levels of engagement across the two conditions.

These differences could arise because the opt-in framing in the Voluntary-Condition may have portrayed the program as optional rather than a job requirement (Hossain and List, 2012; Hong et al., 2015). However, meeting rates are still relatively high even in the Voluntary-Condition, suggesting that differences in meeting rates alone are unlikely to explain the full remaining gap in treatment effects across conditions.

We therefore attempt to test whether the quality of meetings differ across conditions using worksheet contents. Using two approaches, we only find minimal differences. First, we consider the amount of content transcribed on the mentor-protégé worksheets by counting the total number of words written. While this is an imperfect measure of the quality of the mentor-protégé meetings, it proxies for the agents' level of engagement. In our second

[37] While the pre-registered mentoring protocol called for one meeting per week for four weeks, there were instances in which either a mentor, protégé, or both were absent from work for an extended period of time (e.g., on vacation), reducing the number of possible scheduled meetings from four to three (or fewer, in some cases). As such, the denominator of the meeting completion ratio is occasionally less than four.

approach, which is motivated by the worksheet analysis in Sandvik et al. (2020), we use a bag-of-words to determine how much of a response's content is focused on job-specific skills and knowledge, relative to how much is focused on receiving support or encouragement.[38]

In comparing the worksheet content of Mandatory-Condition agents and Voluntary-Condition agents (reported in Table I.A.9), we do not find statistically significant differences in the number of total words or words related to sales skills or knowledge. Mandatory-Condition agents do use about 0.13 more support words than do Voluntary-Condition agents, but this effect is only marginally significant. To the extent that support words signal engagement or encouragement around both the program and the work, this evidence is mildly supportive of the framing channel. However, we have no direct evidence that this interpretation is valid, and we caution that these worksheets are an incomplete record of sentiment or buy-in.[39]

4.4.2 Other Alternative Explanations

A number of other factors, like SUTVA violations, crowd-out, retention differences, or perceptions of differential treatment could potentially explain differences between the Mandatory- and Voluntary-Conditions. We do not find evidence for these alternative explanations. As such, for brevity, we discuss these explanations and their associated tests in detail in Section I.A, and we only discuss the conclusions of the tests here. We refer interested readers to the appendix material for additional detail. In Section I.A.1, we do not find evidence that our results are driven by experimenter demand effects, Hawthorne effects, discouragement from treatment status, or information leakage. Leveraging variation from hold-out cohorts

[38]Specifically, we tabulate the number of "skill" words an agent uses in their responses, and we do the same thing for the number of "support" words. Words that are not classified as either support words or skill words are categorized as "other," including stop words. We list the words in each category in the Internet Appendix (see Section I.E), along with multiple example responses.

[39]Two weeks after mentors and protégés completed their final meeting, the staff asked protégés to complete a post-mentorship survey. The completion rates for this survey were quite low (less than 10%), as the firm did not monitor or provide incentives for completion. Figure I.A.3 shows that protégés, on average, felt like they benefited from the program. The average respondent reported that mentorship helped them to learn selling tactics and that the program increased their day-to-day satisfaction at work.

suggests that violations of the Stable Unit Treatment Value Assumption (SUTVA) are not a major concern for our design. In Section I.A.2, we show that the program does not appear to crowd out organic mentoring that may have occurred in its absence, as non-treated agents in experimental cohorts had similar productivity to agents entering the firm prior to the program's existence.

In Section 4.2, we showed that, across both conditions, the mentorship program had no impact on retention, which, therefore, cannot explain the observed sales revenue treatment effects. In Section I.A.3, we further show that productivity gains remain (i) when accounting for non-random attrition by filling in missing data after separations with the average productivity of replacements and (ii) when using Lee (2009) bounds estimators. The bounding estimator trims the highest and lowest values of the productivity distribution based on the observed attrition rate in the sample. If attrition is non-random with respect to the underlying sales measures, this exercise captures robust treatment effects that are not driven by differential retention of heterogeneous workers across conditions. The bounded estimates of mentorship are never positive for the Voluntary-Condition and remain positive and statistically significant for the Mandatory-Condition.

We show in Section I.A.4 that perceptions of differential treatment, where agents may have bought in more if they felt special for receiving a mentor, are not likely at play. Then we conclude in Section I.A.5 with a discussion of the challenges in directly calibrating belief- or preference-based explanations for agents' opt-in or opt-out behavior. As such, in the next section (Section 5), we rely on indirect measures from national surveys that provide context for belief and preference heterogeneity to determine participation decisions.

5 Returns to the Program, Program Prevalence, and External Validity

In this section, we value the program for the firm and discuss the costs of misallocating mentors to agents with relatively small treatment gains. Then we discuss the results of a nationally representative survey that we conducted, highlighting the widespread prevalence and design variation of human capital development programs. To finish, we comment on the external validity of our findings.

5.1 Net Present Value of Mandatory Mentorship and the Costs of Misallocation

The net present value of the mandatory mentorship program to the firm is equal to approximately \$439,000. To arrive at this estimate, we calculate additional revenues of approximately \$536,000 in present discounted value over a six-month period. The revenue estimates come from an analysis that considers the additional revenue gain to the firm from each investment in mentorship (or each program slot). We track workers over an entire six month period after being allocated to receive a mentor or to the control group. If the worker leaves the firm prior to the six-month horizon, we account for the productivity of replacements by filling in a random draw from the distribution of new agents. In this way, we track both long-term revenue gains and any potential impact on retention from the program. We then subtract \$97,000 of costs, which include costs of the firm's staff to administer the program and each mentor's opportunity cost of lost revenue from engaging in meetings rather than answering calls. We provide details about the calculations Section I.F of the Internet Appendix.

Had the firm allocated all workers to the Mandatory-Condition and had the treatment effects been the same across workers, the firm would have gained an additional \$620,000. If, instead, only about one-third of the treatment gains are due to selection and heterogeneous treatment effects (what our back-of-the-envelope calculation based on the opt-out propensity score recovers), the firm still would have gained approximately \$207,000 in additional revenue had the mentorship treatment been mandatory.

5.2 Prevalence of Human Capital Development Programs

Beyond our study firm, we conducted a nationally representative worker survey to provide background context about human capital development programs, with a focus on three questions: how prevalent are they, how is their participation determined (i.e., mandatory or voluntary), and how often do workers participate when programs are voluntary?

We administered the survey through the Lucid platform in June of 2022 and compensated respondents between \$1 and \$4. The survey took between 7–10 minutes to complete. Respondents had to be employed and pass attention checks to proceed through the survey. We asked respondents whether their current employer offers the following programs: (i) mentorship, (ii) training for new hires, and (iii) ongoing training or continuing education. We also asked whether the programs were required/mandatory or optional/voluntary and, if voluntary, whether they participated. We then probed for the reasons for their participation decisions. We present the results from this survey and details about the survey instrument in Table 7.[40]

The survey responses provided three main takeaways: (1) human capital development programs are ubiquitous; (2) many are voluntary; and (3) many employees do not participate in voluntary programs. Specifically, 45% of the respondents said their employer offers a mentorship program, 87% said they offer new hire training, and 80% said they offer ongoing training or continuing education. About 59% of the mentorship programs and 43% of the continuing education programs offered are voluntary. New hire training is much more likely than the other programs to be mandatory. The last column in Table 7 shows substantial non-participation rates in voluntary programs. Roughly 27%–28% of respondents did not participate in their employer's voluntary mentorship or ongoing training/continuing education programs. Even for new hire training, rates of non-participation exceed 20% when training is optional. Time commitments and doubts about personal program benefits are the most common reasons workers cite for their lack of take-up.[41] These survey results highlight

[40] We also included workplace wellness programs as a validation check. Sixty-five percent of our respondents indicated that their workplace has a wellness program. This is roughly comparable to numbers cited by Jones et al. (2019) from a 2016 Kaiser Family Foundation report, indicating that 53% of firms with more than 200 employees do biometric screening, 59% assess lifestyle health habits, and 83% have programs that encourage healthy lifestyles.

[41] Forty-seven percent of non-participants in mentorship, 36% in new hire training, and 42% in ongoing training cite time constraints or the inconvenience of program offerings as one of their reasons for not participating. "Didn't believe these programs would benefit me" (26% for mentorship, 28% for new hire

the importance of considering the implications of the mandatory versus voluntary participation design choice that many mangers are faced with when they implement a new human capital development program.

5.3 External validity

As part of the first wave of evidence on mandatory versus voluntary programs, we made multiple decisions to give us high internal validity (List, 2020). The tasks that agents performed in the mentorship program—reflecting on their work, sharing these thoughts with mentors, and acting on their mentors' advice—were a natural extension of their day-to-day activities. Our intervention intentionally included features which would allow the treatment to be deployed at scale (permanently) both at the focal firm and in organizations more broadly.

Several additional points suggest our results are likely to be externally valid for workers in other frontline or entry-level jobs. In particular, in data from the Census Bureau's 2019 American Community Survey, sales and related occupations are the second most common entry level job for workers under 25 years old, following food services occupations. While our results may not speak to human capital development programs for stable, professional occupations, our results are applicable to the development decisions of firms that onboard and recruit a substantive share of domestic, entry-level jobs. In addition, our representative worker survey found substantial rates of non-participation in human capital development programs, suggesting that non-participation is a general phenomenon that applies beyond entry-level occupations, as older workers and those with a bachelors degree or higher have slightly higher rates of non-participation in voluntary programs than the overall sample average.

training, and 31% for ongoing training) is the next most common reason. Other options such as, "Didn't plan to stay at the firm, so didn't invest," "Wanted to avoid interaction with coworkers or bosses," and "Felt the program would benefit my employer more than it would benefit me" were selected by 8%–13% of the respondents."

6 Conclusion

Many firms make considerable investments in human capital development programs, such as training and mentorship programs. But an important and understudied question is whether human capital development resources are allocated to the right workers. We consider a ubiquitous decision that managers face when deciding how to allocate human capital development resources: whether they should make development programs mandatory or voluntary. We investigate the implications of this mandatory versus voluntary design choice by conducting a field experiment on mentorship in a U.S.-based inbound sales call center.

We find that a mandatory version of the mentorship program significantly raised workers' productivity, with average sales gains on the order of about 19% over new hires' first two months on the job. By contrast, treatment gains were approximately zero when the program was voluntary. A substantial part of the difference in the efficacy of the mandatory and voluntary programs arises because program treatment effects are negatively correlated with the propensity to participate in the program. Our findings indicate that the decision to make a human capital development program mandatory versus voluntary is not trivial, as the returns to the program are largely determined by selection and treatment effects. As such, our findings shed additional light on why wage inequality and performance differences may persist across workers and firms. That these differences exist even in the presence of high-powered incentive pay suggests that managers may need to mandate worker participation in human capital development programs.

In our setting, training that leverages know-how from coworkers can improve the productivity of lower-performing workers, but low-ability workers may be the *least* likely to seek out such help. As such, an employee's voluntary decision to participate, or not, in human capital development programs may be a useful signal to managers—in this setting and possibly others—of who will benefit the most from additional help. Other allocation rules may also be feasible, but these likely entail waiting to collect performance data, and subsequently staging a performance improvement intervention. In high turnover frontline jobs, firms face a trade-off between a delay in upskilling workers to improve program allocations through targeting versus offering broader training more quickly.

In general, frictions around program participation deserve further investigation, since non-participation is a pervasive feature found in the focal organization and, as shown in our national survey, in other public and private firms. Selection bias in program recruitment can distort program efficacy and inferences, as demonstrated by negative sorting on gains in charter school enrollment (Walters, 2018) and the site selection bias identified by Allcott (2015). With remote work, these allocation questions may become more pronounced (Bojinov et al., 2021; Emanuel et al., 2023), making human capital development allocation decisions even more salient and challenging. Furthermore, the implications of the mandatory versus voluntary program design choice may vary depending on other features of the setting (e.g., how well-defined the content of the training is). These considerations should motivate future research to understand how to allocate scarce human capital development resources.

References

Acemoglu, Daron, Jörn-Steffen Pischke. 1999. The structure of wages and investment in general training. *Journal of Political Economy* **107**(3) 539–572.

Allcott, Hunt. 2015. Site selection bias in program evaluation. *The Quarterly Journal of Economics* **130**(3) 1117–1165.

Allen, Tammy D, Lillian T Eby, Georgia T Chao, Talya N Bauer. 2017. Taking stock of two relational aspects of organizational life: Tracing the history and shaping the future of socialization and mentoring research. *Journal of Applied Psychology* **102**(3) 324.

Anderson, Michael 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) 1481–1495.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, Luigi Zingales. 2024. Meaning at work .

Bandiera, Oriana, Iwan Barankay, Imran Rasul. 2013. Team incentives: Evidence from a firm-level experiment. *Journal of the European Economic Association* **11**(5) 1079–1114.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, Raffaella Sadun. 2020. Ceo behavior and firm performance. *Journal of Political Economy* **128**(4) 1325–1369.

Becker, Gary S. 1975. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Benmelech, Efraim, Nittai Bergman, Hyunseob Kim. 2022. Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources* (S200-S250).

Benson, Alan, Danielle Li, Kelly Shue. 2019. Promotions and the peter principle. *The Quarterly Journal of Economics* **134**(4) 2085–2134.

Benson, Alan, Danielle Li, Kelly Shue. 2024. Potential and the gender promotions gap. *Available at SSRN* .

Bertrand, Marianne, Antoinette Schoar. 2003. Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics* **118**(4) 1169–1208.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, John Van Reenen. 2019. What drives differences in management practices? *American Economic Review* **109**(5) 1648–83.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, John Van Reenen. 2014. Jeea-fbbva lecture 2013: The new empirical economics of management. *Journal of the European Economic Association* **12**(4) 835–876.

Bloom, Nicholas, James Liang, John Roberts, Zhichun Jenny Ying. 2015. Does working from home work? evidence from a chinese experiment. *The Quarterly journal of economics* **130**(1) 165–218.

Bloom, Nicholas, Raffaella Sadun, John Van Reenen. 2016. Management as a technology? Tech. rep., National Bureau of Economic Research.

Bloom, Nicholas, John Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* **122**(4) 1351–1408.

Bojinov, Iavor, Prithwiraj Choudhury, Jacqueline N Lane. 2021. Virtual watercoolers: A field experiment on virtual synchronous interactions and performance of organizational newcomers. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (21-125).

Brynjolfsson, Erik, Danielle Li, Lindsey R Raymond. 2023. Generative ai at work. Tech. rep., National Bureau of Economic Research.

Carrell, Scott E., Bruce I. Sacerdote, James E. West. 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica* **81**(3) 855–882.

Carter, Susan Payne, Whitney Dudley, David S Lyle, John Z Smith. 2019. Who's the boss? The effect of strong leadership on employee turnover. *Journal of Economic Behavior & Organization* **159** 323–343.

Castro, Silvia, Florian Englmaier, Maria Guadalupe. 2022. Fostering psychological safety in teams: Evidence from an rct .

Chan, David C, Matthew Gentzkow, Chuan Yu. 2022. Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics* **137**(2) 729–783.

Chandrasekhar, Arun G, Benjamin Golub, He Yang. 2018. Signaling, shame, and silence in social learning. Tech. rep., National Bureau of Economic Research.

Chatterji, Aaron, Solène Delecourt, Sharique Hasan, Rembrand Koning. 2019. When does advice impact startup performance? *Strategic Management Journal* **40**(3) 331–356.

Chetty, Raj, John N Friedman, Jonah E Rockoff. 2014. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review* **104**(9) 2633–79.

Cornelissen, Thomas, Christian Dustmann, Uta Schönberg. 2017. Peer effects in the workplace. *American Economic Review* **107**(2) 425–456.

Cornwell, Christopher, Ian M Schmutte, Daniela Scur. 2021. Building a productive workforce: The role of structured management practices. *Management Science* **67**(12) 7308–7321.

Coviello, Decio, Andrea Ichino, Nicola Persico. 2014. Time allocation and task juggling. *American Economic Review* **104**(2) 609–23.

Currie, Janet, W Bentley MacLeod. 2017. Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of Labor Economics* **35**(1) 1–43.

Currie, Janet M, W Bentley MacLeod. 2020. Understanding doctor decision making: The case of depression treatment. *Econometrica* **88**(3) 847–878.

Custódio, Cláudia, Miguel A Ferreira, Pedro Matos. 2019. Do general managerial skills spur innovation? *Management Science* **65**(2) 459–476.

Dahl, Gordon B, Katrine V Løken, Magne Mogstad. 2014. Peer effects in program participation. *American Economic Review* **104**(7) 2049–2074.

Del Carpio, Lucia, Maria Guadalupe. 2022. More women in tech? Evidence from a field experiment addressing social identity. *Management Science* **68**(5) 3196–3218.

Delfino, Alexia, Andrea Garnero, Sergio Inferrera, Marco Leonardi, Raffaella Sadun. 2024. Unwilling to reskill? Evidence from a survey experiment with jobseekers.

DellaVigna, Stefano, John A. List, Ulrike Malmendier. 2012. Testing for Altruism and Social Pressure in Charitable Giving *. *The Quarterly Journal of Economics* **127**(1) 1–56. doi:10.1093/qje/qjr050. URL https://doi.org/10.1093/qje/qjr050.

Deming, David J, Justine S Hastings, Thomas J Kane, Douglas O Staiger. 2014. School choice, school quality, and postsecondary attainment. *American Economic Review* **104**(3) 991–1013.

Diaz, Brayan, Andrea Neyra-Nazarrett, Julian Ramirez, Raffaella Sadun, Jorge Tamayo. 2024. Training within firms .

Edmans, Alex. 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics* **101**(3) 621–640.

Edmondson, Amy C., Zhike Lei. 2014. Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior* **1**(1) 23–43. doi:10.1146/annurev-orgpsych-031413-091305.

Emanuel, Natalia, Emma Harrington, Amanda Pallais. 2023. The power of proximity to coworkers: training for tomorrow or productivity today? Tech. rep., National Bureau of Economic Research.

Englmaier, Florian, Nicolai J Foss, Thorbjørn Knudsen, Tobias Kretschmer. 2018. Organization design and firm heterogeneity: Towards an integrated research agenda for strategy. *Organization design* 229–252.

Englmaier, Florian, Stefan Grimm, Dominik Grothe, David Schindler, Simeon Schudy. 2021. The value of leadership: Evidence from a large-scale field experiment .

Espinosa, Miguel, Christopher Stanton. 2021. Worker skills and organizational spillovers: Evidence from linked training and communications data. Tech. rep., Harvard Business School.

Finkelstein, Amy, Matthew Gentzkow, Heidi Williams. 2016. Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics* **131**(4) 1681–1726.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, Nick Zubanov. 2023. What do employee referral programs do? measuring the direct and overall effects of a management practice. *Journal of Political Economy* **131**(3) 633–686.

Friebel, Guido, Matthias Heinz, Miriam Krueger, Nikolay Zubanov. 2017. Team incentives and performance: Evidence from a retail chain. *American Economic Review* **107**(8) 2168–2203.

Friebel, Guido, Matthias Heinz, Nikolay Zubanov. 2022. Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain. *Management Science* **68**(1) 211–229.

Fudenberg, Drew, Luis Rayo. 2019. Training and effort dynamics in apprenticeship. *American Economic Review* **109**(11) 3780–3812.

Gibbons, Robert, Rebecca Henderson. 2012. *What do managers do? Exploring persistent performance differences among seemingly similar enterprises*. Harvard Business School.

Ginther, Donna K, Janet M Currie, Francine D Blau, Rachel TA Croson. 2020. Can mentoring help female assistant professors in economics? An evaluation by randomized trial. *AEA Papers and Proceedings*, vol. 110. 205–09.

Gosnell, Greer K, John A List, Robert D Metcalfe. 2020. The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy* **128**(4) 1195–1233.

Gubler, Timothy, Ian Larkin, Lamar Pierce. 2018. Doing well by making well: The impact of corporate wellness programs on employee productivity. *Management Science* **64**(11) 4967–4987.

Gutner, Toddi. 2009. Finding anchors in the storm: Mentors. *The Wall Street Journal* .

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

Heckman, James J, Hidehiko Ichimura, Petra E Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* **64**(4) 605–654.

Herbst, Daniel, Alexandre Mas. 2015. Peer effects on worker output in the laboratory generalize to the field. *Science* **350**(6260) 545–549.

Heursen, Lea, Svenja Friess, Marina Chugunova. 2023. Reputational concerns and advice-seeking at work .

Hilmer, Christiana, Michael Hilmer. 2007. Women helping women, men helping women? Same-gender mentoring, initial job placements, and early career publishing success for economics phds. *The American economic review* **97**(2) 422–426.

Hoffman, Mitchell, Stephen V Burks. 2020. Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics* **11**(1) 315–348.

Hoffman, Mitchell, Lisa B. Kahn, Danielle Li. 2017. Discretion in hiring. *The Quarterly Journal of Economics* **133**(2) 765–800.

Hoffman, Mitchell, Christopher T Stanton. 2024. People, practices, and productivity: A review of new advances in personnel economics .

Hoffman, Mitchell, Steven Tadelis. 2021. People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy* **129**(1) 000–000.

Hong, Fuhai, Tanjim Hossain, John A List. 2015. Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization* **118** 372–382.

Hossain, Tanjim, John A List. 2012. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science* **58**(12) 2151–2167.

Huffman, David, Collin Raymond, Julia Shvets. 2022. Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review* **112**(10) 3141–75.

Johnson, Matthew S., David I. Levine, Michael W. Toffel. 2023. Improving regulatory effectiveness through better targeting: Evidence from osha. *American Economic Journal: Applied Economics* **15**(4) 30–67. doi:10.1257/app.20200659. URL https://www.aeaweb.org/articles?id=10.1257/app.20200659.

Jones, Damon, David Molitor, Julian Reif. 2019. What do workplace wellness programs do? Evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics* **134**(4) 1747–1791.

Karlan, Dean, Jonathan Zinman. 2009. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica* **77**(6) 1993–2008.

Larkin, Ian. 2014. The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics* **32**(2) 199–227.

Larkin, Ian, Stephen Leider. 2012. Incentive schemes, sorting and behavorial biases of employees: Experimental evidence. *American Economic Journal: Microeconomics* **4**(2) 184–214.

Larrain, Mauricio. 2015. Capital account opening and wage inequality. *The Review of Financial Studies* **28**(6) 1555–1587.

Lazear, Edward P., Ulrike Malmendier, Roberto A. Weber. 2012. Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics* **4**(1) 136–63. doi:10.1257/app.4.1.136. URL `https://www.aeaweb.org/articles?id=10.1257/app.4.1.136`.

Lazear, Edward P, Kathryn L Shaw, Christopher Stanton. 2016. Making do with less: working harder during recessions. *Journal of Labor Economics* **34**(S1) S333–S360.

Lazear, Edward P, Kathryn L Shaw, Christopher T Stanton. 2015. The value of bosses. *Journal of Labor Economics* **33**(4) 823–861.

Lee, David S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* **76**(3) 1071–1102.

Li, Danielle, Lindsey R Raymond, Peter Bergman. 2020. Hiring as exploration. Tech. rep., National Bureau of Economic Research.

List, John A. 2020. Non est disputandum de generalizability? A glimpse into the external validity trial. Tech. rep., National Bureau of Economic Research.

List, John A. 2022. *The voltage effect: How to make good ideas great and great ideas scale*. Currency.

Lyle, David S, John Z Smith. 2014. The effect of high-performing mentors on junior officer promotion in the us army. *Journal of Labor Economics* **32**(2) 229–258.

Mas, Alexandre, Enrico Moretti. 2009. Peers at work. *American Economic Review* **99**(1) 112–45.

Metcalfe, Robert D, Alexandre B Sollaci, Chad Syverson. 2023. Managers and productivity in retail. Tech. rep., National Bureau of Economic Research.

Minni, Virgnia. 2023. Making the invisible hand visible: Managers and the allocation of workers to jobs. *Working paper, Chicago Booth* .

Nishesh, Naman, Paige Ouimet, Elena Simintzi. 2022. Labor and corporate finance. *Available at SSRN* .

Oyer, Paul. 1998. Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *The Quarterly Journal of Economics* **113**(1) 149–185.

Oyer, Paul, Scott Schaefer. 2011. Personnel economics: Hiring and incentives. *Handbook of Labor Economics* **4** 1769–1823.

Reif, Julian, David Chan, Damon Jones, Laura Payne, David Molitor. 2020. Effects of a workplace wellness program on employee health, health beliefs, and medical use: a randomized clinical trial. *JAMA Internal Medicine* **180**(7) 952–960.

Rockoff, Jonah E. 2008. Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in new york city. Tech. rep., National Bureau of Economic Research.

Rouen, Ethan, Matthias Regier. 2022. The stock market value of human capital creation. *Harvard Business School Accounting & Management Unit Working Paper* (21-047).

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, Christopher T Stanton. 2020. Workplace knowledge flows. *The Quarterly Journal of Economics* **135**(3) 1635–1680.

Starr, Evan. 2019. Consider this: Training, wages, and the enforceability of covenants not to compete. *ILR Review* **72**(4) 783–817.

Statista. 2022. Total training expenditures in the united states from 2012 to 2022 (in billion u.s. dollars) [graph]. `https://www-statista-com.ezp-prod1.hul.harvard.edu/statistics/788521/training-expenditures-united-states/`. Retrieved on 2023-01-29.

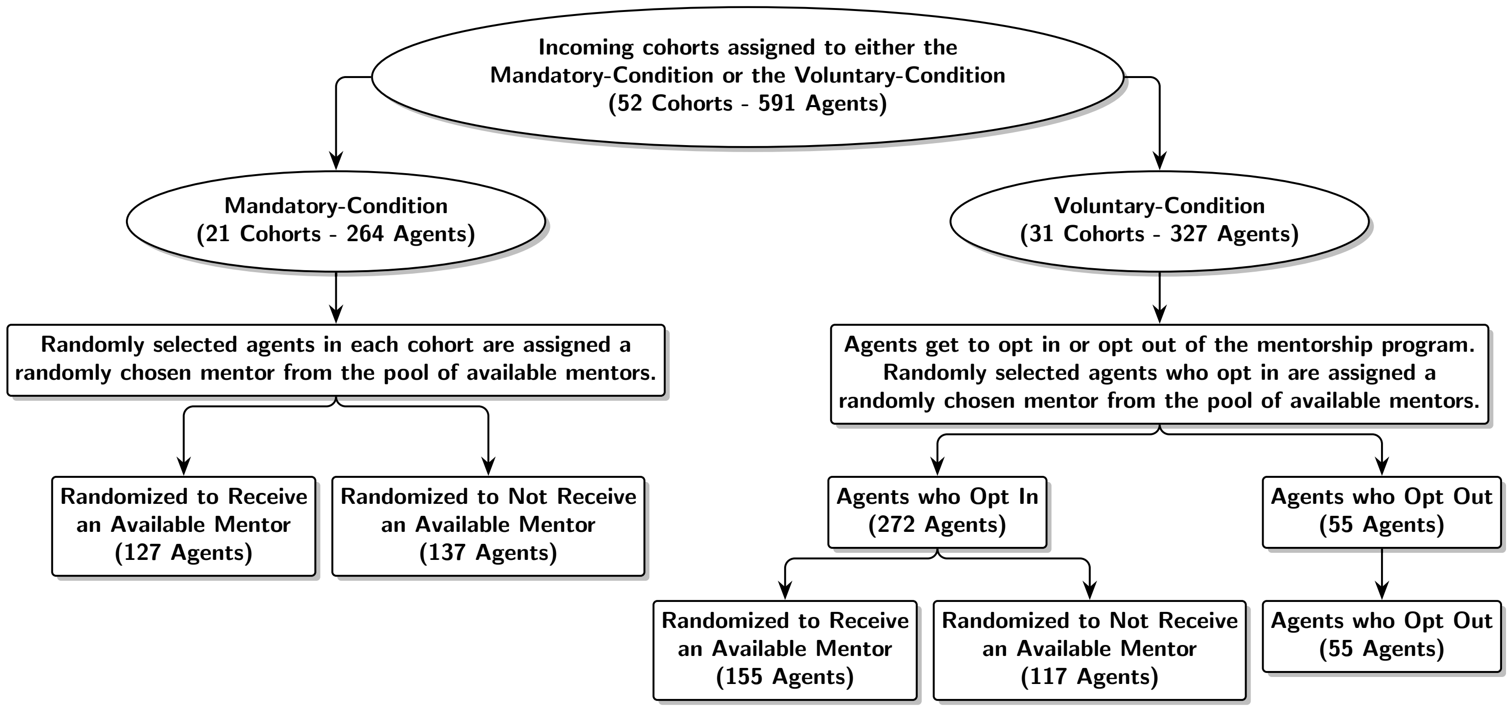
Syverson, Chad. 2011. What determines productivity? *Journal of Economic literature* **49**(2) 326–65.

Walters, Christopher R. 2018. The demand for effective charter schools. *Journal of Political Economy* **126**(6) 2179–2223.

Zingales, Luigi. 2000. In search of new foundations. *The Journal of Finance* **55**(4) 1623–1653.

Zivin, Joshua Graff, Lisa B Kahn, Matthew Neidell. 2021. Incentivizing learning-by-doing: The role of compensation schemes. *Workplace Productivity and Management Practices*. Emerald Publishing Limited, 139–178.

Figure 1: Allocation of Cohorts and Agents to Treatment Conditions



Notes. This figure displays the allocation of the 52 mentor-eligible cohorts to either the Mandatory-Condition or the Voluntary-Condition, our first level of variation. It then shows the allocation of the 591 mentor-eligible agents within these cohorts into different treatment conditions, our second level of variation. This is based on agents who complete training and are observed to have post-training productivity data.

Table 1: Balance Tests for Treatment Assignment

Panel A: Cohort-Level Balance in Agent Characteristics

	Mandatory-Condition	Voluntary-Condition	p-value
	(1)	(2)	(2)−(1)
Age (yrs.)			
Mean	22.70	22.80	0.887
Std Dev.	(2.40)	(2.34)	
Female			
Mean	0.43	0.40	0.624
Married			
Mean	0.13	0.16	0.522
Hiring Score			
Mean	0.83	0.85	0.207
Std Dev.	(0.04)	(0.04)	
Adjusted Hiring Score			
Mean	0.84	0.86	0.029
Std Dev.	(0.03)	(0.03)	
Referral			
Mean	0.57	0.58	0.746
N Cohorts	21	31	

Panel B: Balance in Agent Characteristics For Those Eligible for Mentor Assignment

	Mandatory-Condition			Voluntary-Condition		
	Mentored	Non-Mentored	p-value	Mentored	Non-Mentored	p-value
	(1)	(2)	(2)−(1)	(3)	(4)	(4)−(3)
Age (yrs.)						
Mean	22.40	23.51	0.193	22.47	22.51	0.945
Std Dev.	(4.46)	(8.60)		(5.54)	(6.18)	
Female						
Mean	0.46	0.40	0.303	0.45	0.38	0.318
Married						
Mean	0.09	0.15	0.150	0.15	0.17	0.722
Hiring Score						
Mean	0.83	0.84	0.432	0.85	0.86	0.508
Std Dev.	(0.09)	(0.08)		(0.08)	(0.08)	
Adj. Hiring Score						
Mean	0.83	0.84	0.322	0.86	0.86	0.521
Std Dev.	(0.08)	(0.07)		(0.07)	(0.07)	
Referral						
Mean	0.58	0.55	0.649	0.56	0.60	0.543
Number of Agents	127	137		155	117	

Notes. This table presents balance tests. Most characteristics are self-explanatory other than the Hiring Score, which is a recruiter-assigned measure of fit with the job, ranging from 0 to 1. The Adjusted (Adj.) Hiring Score accounts for individual recruiter leniency, estimated using the productivity of non-mentor-eligible agents outside of the experiment, as described in Section 3.5. In Panel A, we report average agent characteristics at the cohort-level to test for assignment balance between the Mandatory- and Voluntary-Conditions. In Panel B, we test for balance in sub-treatment assignment to mentors. In Panel B, the Voluntary-Condition sample is not comparable to the Mandatory-Condition sample due to selection into the program (Table 4 compares the characteristics of those who opt in and opt out in the Voluntary-Condition). Standard deviations are in parentheses for continuous variables. The p-values come from difference-in-means tests across high-level treatment conditions in Panel A and for agents who do and do not receive mentors among those eligible for assignment in Panel B.

Table 2: Treatment and Selection Effects of Mentoring on Productivity

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.171***	0.112***	-0.084	-0.084			0.171***	0.112***	0.145***
standard errors	(0.039)	(0.027)	(0.080)	(0.054)			(0.038)	(0.026)	(0.052)
sharpened q-value							[0.001]	[0.002]	[0.013]
Voluntary Opt-Out					-0.369***	-0.264***	-0.277***	-0.161***	-0.141***
standard errors					(0.109)	(0.068)	(0.098)	(0.053)	(0.028)
sharpened q-value							[0.013]	[0.010]	[0.001]
Mentored × Voluntary							-0.272***	-0.207***	-0.203***
standard errors							(0.086)	(0.062)	(0.054)
sharpened q-value							[0.009]	[0.007]	[0.003]
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.028	0.032	0.020	0.046	0.047	0.066	0.026	0.040	0.059
Observations	6,725	6,725	7,569	7,569	4,734	4,734	15,670	15,670	15,670
p-value: Mentored + Mentored × Voluntary							0.241	0.124	0.155

Notes. This table reports estimates of the different treatment and selection effects from the mentorship program. The sample is composed of agent-day productivity data across agents' first two months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include all agents in the Mandatory-Condition. The specifications in Columns (3) and (4) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include all agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition (see the text for additional details). We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. Sharpened q-values that adjust for the false discovery rate are presented in brackets, following Anderson (2008). The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Treatment Effects on Retention

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)	
	Tenure_{30}	Tenure_{60}	Tenure_{30}	Tenure_{60}	Tenure_{30}	Tenure_{60}	Tenure_{30}	Tenure_{60}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mentored	-0.001	-0.009	-0.001	-0.072			-0.001	-0.009
	(0.036)	(0.112)	(0.046)	(0.060)			(0.035)	(0.108)
Voluntary Opt-Out					-0.117	-0.190	-0.054	-0.196*
					(0.079)	(0.126)	(0.058)	(0.104)
Mentored × Voluntary							-0.018	-0.072
							(0.062)	(0.147)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.041	0.009	0.003	0.066	0.036	0.084	0.034	0.040
Observations	264	264	272	272	172	172	591	591
Mean Value of Tenure_i	0.86	0.61	0.92	0.67	0.91	0.65	0.89	0.63
p-value: Mentored + Mentored × Voluntary							0.651	0.194

Notes. The sample used is composed of a single observation per agent, among all mentor-eligible agents with post-training productivity data. $Tenure_{30}$ ($Tenure_{60}$) equals one for agents who remain with the firm for at least thirty (sixty) days after their hire date, and zero otherwise. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include all agents in the Mandatory-Condition. The specifications in Columns (3) and (4) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7) and (8) include agents from both conditions. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. The penultimate row reports the average value of $Tenure_i$ for the sample of agents used in the specification within that column. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored × Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Determinants of Program Opt-Out and the Relationship Between Opting Out, Productivity, and Worker Characteristics

Dep. Variable	= 1 if Opted Out				IHS(Revenue)		IHS(RPC)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.002	0.001	0.002	-0.001	-0.002	-0.004	-0.001
	(0.002)	(0.002)	(0.001)	(0.003)	(0.005)	(0.005)	(0.003)
Female	-0.064**	-0.011	-0.007	-0.019	-0.104	-0.093	-0.084
	(0.029)	(0.030)	(0.032)	(0.032)	(0.136)	(0.128)	(0.084)
Married	-0.012	-0.019	-0.017	0.004	0.067	0.119	0.105
	(0.049)	(0.058)	(0.055)	(0.049)	(0.156)	(0.185)	(0.124)
Adjusted Hiring Score	-0.693***	-0.666***	-0.669***	-0.664***	2.703***	2.931***	2.088***
	(0.245)	(0.211)	(0.195)	(0.182)	(0.852)	(0.818)	(0.515)
Location 1	-0.051	0.046	0.037	0.026			
	(0.078)	(0.053)	(0.053)	(0.052)			
Referral	-0.023	0.019	0.016	0.011	-0.188	-0.183	-0.079
	(0.037)	(0.048)	(0.051)	(0.043)	(0.120)	(0.112)	(0.062)
Call Center Exp.		0.093	0.093*	0.099*		0.397**	0.313**
		(0.060)	(0.052)	(0.052)		(0.187)	(0.112)
Sales Experience		0.006	0.006	0.005		-0.017	-0.050
		(0.059)	(0.056)	(0.057)		(0.223)	(0.123)
High Extroversion			-0.025	-0.025		0.205	0.163
			(0.041)	(0.042)		(0.131)	(0.098)
High Agreeableness			-0.040	-0.040		-0.206*	-0.109
			(0.026)	(0.025)		(0.106)	(0.088)
High Conscientiousness			-0.059	-0.056		-0.035	-0.051
			(0.047)	(0.045)		(0.110)	(0.076)
High Emotional Stability			0.042	0.041		-0.146	-0.056
			(0.052)	(0.053)		(0.099)	(0.054)
High Openness			0.024	0.027		0.039	0.062
			(0.037)	(0.037)		(0.138)	(0.088)
Missing Survey				0.288***		-0.457**	-0.298**
				(0.052)		(0.180)	(0.121)
Division Fixed Effects		✓	✓	✓			
Cohort Fixed Effects					✓	✓	✓
(Pse.) R-Square	0.036	0.070	0.085	0.209	0.061	0.073	0.094
Observations	365	341	341	365	4,734	4,734	4,734

Notes. The sample in Columns (1)–(4) is restricted to the 365 agents in the Voluntary-Condition, including those who quit before they completed training. The dependent variable is an indicator that equals one if the agent opted out of the program. The coefficients capture the marginal effects of a unit change in each regressor from logistic regressions of different predictors of the choice to opt out. Experience and personality factors were collected via survey. We split personality scores on the sample median. Column (4) includes agents who did not complete the new hire survey, which we account for with a *Missing Survey* indicator. In Columns (5)–(6), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(Revenue), the inverse hyperbolic sine of daily revenue, on agents' characteristics. In Column (7), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(RPC), the inverse hyperbolic sine of daily revenue-per-call, on agents' characteristics. In Columns (2)–(4), the Division Fixed Effects indicators reflect the inclusion of controls for whether the agent works in the first largest division, the second largest division, or one of the other smaller divisions. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level). We report marginal effects and delta-method standard errors in parentheses in Columns (1)–(4). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Treatment Effects of Mentoring in the Mandatory-Condition by Predicted Opt-Out Propensity

	IHS(Revenue)			IHS(RPC)		
	High_{Opt}	Low_{Opt}	All	High_{Opt}	Low_{Opt}	All
	(1)	(2)	(3)	(4)	(5)	(6)
Mentored	0.324**	0.069**	0.063*	0.262**	0.026	0.022
	(0.138)	(0.023)	(0.029)	(0.094)	(0.021)	(0.021)
Mentored × High_{Opt}			0.342**			0.285***
			(0.128)			(0.078)
High_{Opt}			-0.239**			-0.163***
			(0.085)			(0.049)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.038	0.036	0.030	0.044	0.035	0.035
Observations	2,244	4,481	6,725	2,244	4,481	6,725

Notes. This table reports heterogeneous treatment effect estimates for agents in the Mandatory-Condition. We estimate agents' opt-out propensity scores as described in Section 4.3.2. After estimating propensity scores, we place agents into $High_{Opt}$ if their propensity score of opting out is in the top 33.3% of the propensity score distribution, and we place agents with a propensity score in the bottom 66.7% into Low_{Opt}, indicating that they had a low likelihood to opt out. We use a larger threshold than the opt-out rate in the Voluntary-Condition because (i) the individual propensity scores are less than 1, implying we need more workers to approximate the total number of those who opt out in the Voluntary-Condition and (ii) we want a sample that is large enough for reliable inference. We then estimate Equation (1) within these subsets of the data with either IHS(Revenue) or IHS(RPC) as the dependent variable. To determine if the effect of mentorship is significantly different between the $High_{Opt}$ and Low_{Opt} agents, we pool the samples in Columns (3) and (6) and include a one-zero indicator for $High_{Opt}$ along with its interaction with *Mentored*. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Meeting Completion Rates Across Conditions in the Experiment

	Mandatory-Condition	Voluntary-Condition	Diff. p-value	Mandatory Low_{Opt}	Diff. p-value
	(1)	(2)	(3)	(4)	(5)
Number of Agents	127	155		89	
At Least One Recorded Meeting	109	130		77	
No Recorded Meeting	18	25		12	
Number Recorded Meetings (avg.)	2.31	2.11	0.260	2.42	0.115
	(1.58)	(1.36)		(1.61)	
Meeting Completion Ratio (avg.)	0.74	0.64	0.031	0.77	0.014
	(0.38)	(0.38)		(0.36)	

Notes. In this table we report the mentor meeting completion details of protégés in the Mandatory-Condition, the Voluntary-Condition, and among protégés in the Mandatory-Condition with low opt-out propensity scores. *No Recorded Meeting* indicates that there is no record that the mentor-protégé pair ever met with one another. The *Meeting Completion Ratio* measure is based on the number of possible meetings the mentor-protégé pair could have had. While the pre-registered mentoring protocol called for one meeting per week for four weeks, there were instances in which either a mentor or protégé or both were absent from work for an extended period of time (e.g., on vacation), reducing the number of possible scheduled meetings from four to three (or fewer, in some cases). As such, the denominator of the meeting completion ratio is occasionally less than four. Column (4) considers agents in the Mandatory-Condition with opt-out propensity scores in the bottom two terciles. The p-values in Column (3) are from difference-in-means comparisons of the values in Columns (1) and (2). The p-values in Column (5) are from difference-in-means comparisons of the values in Columns (2) and (4).

Table 7: Survey Data on the Characteristics of Human Capital Development Programs and Participation in Voluntary Programs

Program Type:	Is the Program Offered?	If It Is Offered, Is It Voluntary?	If It Is Voluntary, Do You Not Participate?
Formal Mentorship	0.45	0.59	0.27
	(0.01)	(0.01)	(0.02)
New Hire Training	0.87	0.22	0.21
	(0.01)	(0.01)	(0.02)
Ongoing Training or Cont. Ed.	0.80	0.43	0.28
	(0.01)	(0.01)	(0.01)
N = 3,191			

Notes. This table displays summary statistics on the prevalence and administrative choices for different human capital development programs. Means and standard errors (in parentheses) are reported. Data come from a nationally representative online survey conducted through the Lucid platform in June of 2022. The survey was restricted to respondents currently employed by others. Respondents were asked about whether their employer offers a particular program and whether it is voluntary or mandatory with the question: "Consider your current employer. Which of the following programs does your employer offer to you personally? If offered, are you required to participate (required/mandatory) or can you choose to participate or not (optional/voluntary)?" For each program, respondents chose between "Required or Mandatory," "Optional or Voluntary," or "Not offered." For the three core programs—mentorship, new hire training, and continuing education—if a respondent indicated that a program was voluntary, follow-up questions were asked about their participation and the reasons for their lack of participation, if applicable. As reported in the text, the survey also asked about workplace wellness programs to benchmark responses against other sources.

Internet Appendix to "Should Human Capital Development Programs be Mandatory or Voluntary? Evidence from a Field Experiment on Mentorship"

Contents

I.A Alternative Explanations for Treatment Effects

In this section, we evaluate threats to the validity of our treatment effect estimates, and we explore a range of possible alternative explanations for differences in treatment effects between the Mandatory- and Voluntary-Conditions.

I.A.1 Experimenter Demand Effects, Hawthorne Effects, Discouragement, or Information Leakage

Our design is a natural field experiment (Harrison and List, 2004), where the participants never met the researchers, limiting experimenter demand effects (i.e., observer bias). To participants, the mentoring program appeared like a normal work activity. While participants were told that outside researchers were analyzing their survey and productivity data, the mentorship program was framed and experienced as a regular part of the firm's onboarding process. Furthermore, participants were not aware that differences between the Mandatory- and Voluntary-Conditions were the objects of researcher interest.[1] Similarly, Hawthorne effects were not likely in this setting, as sales managers monitor the same performance metrics that we study and provide workers with performance-related feedback. It is unlikely that subject behavior was impacted by the knowledge that outside researchers—with whom the agents never interacted—were tracking their performance. In our prior work with this firm, benign treatments with no productivity impact allowed us to test for Hawthorne and demand effects; we found no evidence of their importance in this setting (Sandvik et al., 2020).

Discouragement and the possible leakage of information due to the design of the program are always important mechanisms to consider in natural field experiments. In our setting, agents who did not receive a mentor may have become discouraged, reducing their performance. Discouragement effects could potentially result in a difference in productivity between mentored and non-mentored agents; we would expect agents who received a mentor to outperform those that did not. In the Voluntary-Condition, where we would expect discouragement effects to be most salient, we find no evidence that agents who opted in and received a mentor outperformed those who did not. In our implementation, we preemptively worked with the company to limit discouragement and information leakage. For example, the protocol called for the staff to privately notify treated workers of their mentor assignment—reducing the salience of unequal treatment and the potential for discouragement among non-mentored workers. We also asked the staff to monitor any complaints or concerns over not being matched to a mentor; no instances of discouragement were ever communicated to the research team. In addition, they never indicated that the content of the mentor-protégé meetings was shared with non-mentored agents.

We formally test the net effect of discouragement and information leakage by comparing the performance of new workers who joined the firm in experimental cohorts with new hires who were not part of the experiment. Under the null of no net discouragement (or encouragement), information leakage, or other SUTVA violations, we would expect the productivity of non-treated workers in experimental cohorts to be equal to the productivity of new hires

[1] Subjects were asked to provide informed consent when responding to the new hire survey. The survey was framed around understanding employees' preferences, work styles, and personality characteristics so that university researchers could help the firm better serve its workforce. The consent protocol did not specify that selection into or out of the mentoring program was the key metric being studied, as this decision was elicited by the firm's staff.

outside of the experiment. To conduct this test, we compare new hires in experimental cohorts with 1) new hires who were in hold-out cohorts during the time of the experiment and 2) new hires who joined the firm prior to the experiment. Our tests examine the relative performance of new hires versus experienced veterans working in the same divisions at the same time, which enables us to make comparisons over time when sales conditions may differ. This approach removes common time series differences across cohorts, including those due to demand fluctuations, and improves the power of our tests.

We estimate the following model using ordinary least squares:

$$\begin{aligned} \mathrm{y}_{i,t} = \alpha &+ \beta_1 \text{New Hire}_i + \beta_2 (\text{New Hire} \times \text{Mandatory})_i \\ &+ \beta_3 (\text{New Hire} \times \text{Voluntary})_i \\ &+ \beta_4 (\text{New Hire} \times \text{Mandatory} \times \text{Mentored})_i \\ &+ \beta_5 (\text{New Hire} \times \text{Voluntary} \times \text{Mentored})_i + \zeta_{j,l,t} + \varepsilon_{i,t}, \end{aligned} \tag{I.A.1}$$

where *New Hire* is an indicator if the agent has tenure of two months or less, *Mandatory* and *Voluntary* are indicators for the Mandatory- and Voluntary-Conditions among program-eligible cohorts, respectively, and *Mentored* is an indicator for those assigned a mentor. $\zeta_{j,l,t}$ captures division-by-location-by-date fixed effects, absorbing fluctuations in call volumes across divisions on particular dates and location-specific shocks that may affect productivity; $\varepsilon_{i,t}$ is an idiosyncratic error term.

Our test of net discouragement/encouragement and information leakage is the joint test that $\beta_2 = \beta_3 = 0$, indicating that the productivity of new hires relative to veterans in the Mandatory- and Voluntary-Conditions is no different than the new hire-to-veteran productivity differences in hold-out cohorts. In the Voluntary-Condition, those who opt out of the program and those who opt in but are not mentored are a non-random group of agents who are pooled together in the $(\textit{New Hire} \times \textit{Voluntary})_i$ group indicator. We therefore conduct another test of the joint null of zero returns to mentorship and zero net discouragement and information leakage that is robust to non-random selection by testing whether $\beta_3 = \beta_5 = 0$. This test asks whether overall cohort-level productivity in the Voluntary-Condition differs from the productivity of hold-out cohorts.

The results in Panel A of Table I.A.5 report estimations of productivity relative to these hold-out cohorts. The negative and statistically significant coefficient on *New Hire* in Column (1) suggests that newly hired agents generate approximately 30% ($= e^{-0.354} - 1$) less daily revenue relative to veterans. The small and insignificant coefficients on *New Hire* $\times$ *Mandatory* and *New Hire* $\times$ *Voluntary* suggest that newly hired non-mentored agents in program-eligible cohorts perform like newly hired agents in hold-out cohorts. We thus fail to detect evidence of discouragement or leakage. In addition, we are unable to reject the null that $\beta_2 = \beta_3 = 0$, indicating that newly hired, non-mentored agents in the Mandatory-Condition performed like newly hired, non-mentored agents in the Voluntary-Condition. The significant coefficient on *New Hire* $\times$ *Mandatory* $\times$ *Mentored* and the insignificant coefficient on *New Hire* $\times$ *Voluntary* $\times$ *Mentored* align with the main treatment effect estimates discussed in Section 4. Column (2) shows that our results are robust when controlling for agent demographic characteristics—age, gender, and marital status—which is important, given that randomization of agents into treatments did not occur for veterans and hold-out cohort agents (we do not have data on referral status or hiring scores for many veteran agents). Columns (3)–(4) repeat this exercise while using IHS(RPC) as the dependent variable. The small, insignificant coefficients on *New Hire* × *Mandatory* and *New Hire* × *Voluntary* in both of the columns further support the notion that discouragement and leakage are unlikely drivers of our estimated mentorship treatment effects. In the last row of Panel A, we cannot reject the null that productivity in the voluntary cohorts equals the productivity of hold-out cohorts.

I.A.2 Possibility of Program Crowd-Out

Formal mentoring programs have the potential to crowd out informal arrangements that fill similar functions. In our context, the program may have crowded out informal mentorship if the mentors were already providing informal guidance but stopped once the program began. To test whether our program crowded out mentoring that would have occurred in its absence, we assess whether non-mentored agents in program-eligible cohorts perform less well than new agents who joined the firm prior to the program. We again use the relative performance difference between new hires and veterans as the basis for comparison. If the mentorship program crowded out organic mentoring, then we would expect non-mentored new hires during the program to perform worse than new hires from prior years. We find no such evidence in Panel B of Table I.A.5. These estimations resemble those in Panel A, but the comparison group identifying the baseline *New Hire* indicator is now new hires who enter the firm prior to the experiment, rather than those from contemporaneous hold-out cohorts.[2] Non-mentored new hires during the time of the program had similar levels of productivity relative to new hires from before the program, suggesting that the program was unlikely to have crowded out organic mentoring.

I.A.3 Retention Effects and Inference About Sales Treatment Effects

In Section 4.2, we showed that, across both conditions, the mentorship program had no impact on retention, which means differential attrition is unlikely to explain the sales revenue treatment effects. We provide additional evidence of this negligible effect in Table I.A.7, which considers retention at longer horizons (i.e., 90–180 days). There are two reasons to probe further on retention. First, retention is consequential for the firm and small differences in retention may change the unit economics of the program. Second, we want to be confident that retention differences across conditions do not explain the treatment effects on sales.

An exercise that helps with both goals is to create what looks like a balanced panel, where we treat a mentorship slot as the unit of analysis and then track what happens to treatment and control agents over a fixed time horizon. This allows us i) to account for the long-term impact of each investment in mentorship and ii) to test whether the results differ from the main analysis, which would indicate that retention differences might drive some of the treatment gains because of imbalance in the panel that happens over time. Specifically, we fill in the productivity of agents who leave the firm with the expected productivity of a replacement for both mentored and non-mentored agents. The total productivity gain to the firm from a mentorship slot is the relative productivity gain of the treated agent while the agent is retained, followed by the productivity of a randomly drawn replacement post-

[2] Contemporaneous cohorts are not a good comparison group because they would be subject to the same limited supply of informal mentors.

separation. A similar approach is used for the productivity of non-mentored controls.[3] When using the mentorship slot as the unit of analysis and filling in replacement productivity, we find results that largely mirror our main results on the unbalanced panel (see Panel A of Table I.A.10). Using the mentorship slot as the unit of analysis, the per-worker benefits to the firm in the Mandatory-Condition remain positive and significant, whereas the analogous firm-benefits are negligible in the Voluntary-Condition. The fact that the estimates are similar suggests that retention differences do not affect our inference about sales productivity treatment effects.

A second approach is to estimate bounds that are robust to non-random attrition, as proposed by Lee (2009). The intuition behind these tests is that one trims the highest and lowest values of the distribution based on the attrition rate from the sample, which gives bounds on the treatment effect for a non-retention related outcome. To the extent that those who are most likely to exit have the lowest values of the productivity variables, this exercise speaks directly to the retention concern. Table I.A.11 reports the upper and lower bounds of the estimated treatment effects on productivity in months 1–2. The results suggest that our estimated treatment effects in the Mandatory-Condition are largely attributable to the intensive margin of agents becoming more productive, and not due to differential retention effects.[4]

I.A.4 Salience of Differential Treatment

Another possibility is that differences in encouragement may have resulted from randomization into treatment, where receiving a mentor may have caused agents to feel special or exceptional because of perceived inequality in access to benefits provided by the firm. These perceptions may be exacerbated if the randomized nature of treatment was less salient to agents in the Mandatory-Condition compared to the Voluntary-Condition.

To test for such encouragement effects, we compare the treatment effects among mentored agents in teams where relatively more or fewer agents were also mentored. Specifically, we compute the fraction of agent-days where an agent may have been aware of differential treatment by working alongside anyone in their hiring cohort who had a different treatment status. If mentored agents felt they were chosen for mentorship for non-random reasons, then they might have changed their effort or buy-in in response to this perception of the firm's commitment to them. If such perceptions existed, we would expect to see larger treatment effects for agents with fewer mentored teammates, as exposure to more mentored teammates would likely moderate these feelings of being exceptional. We show in Table I.A.12 that the treatment effects of mentorship in the Mandatory-Condition do not vary based on whether or not mentored agents had more exposure to teammates from their same hiring cohort who were also mentored. While we cannot completely rule out the possibility that encouragement effects were at play in our setting, our tests do not detect any evidence of their influence.

[3] Specifically, use the productivity of a randomly chosen newly hired non-mentor-eligible agent in the same location-division-year-quarter as the departed agent. We then re-estimate our main intention-to-treat regression models. We estimate ordinary least squares regressions with cohort fixed effects in all columns. We conduct 200 iterations of this estimation procedure to bootstrap the standard errors.

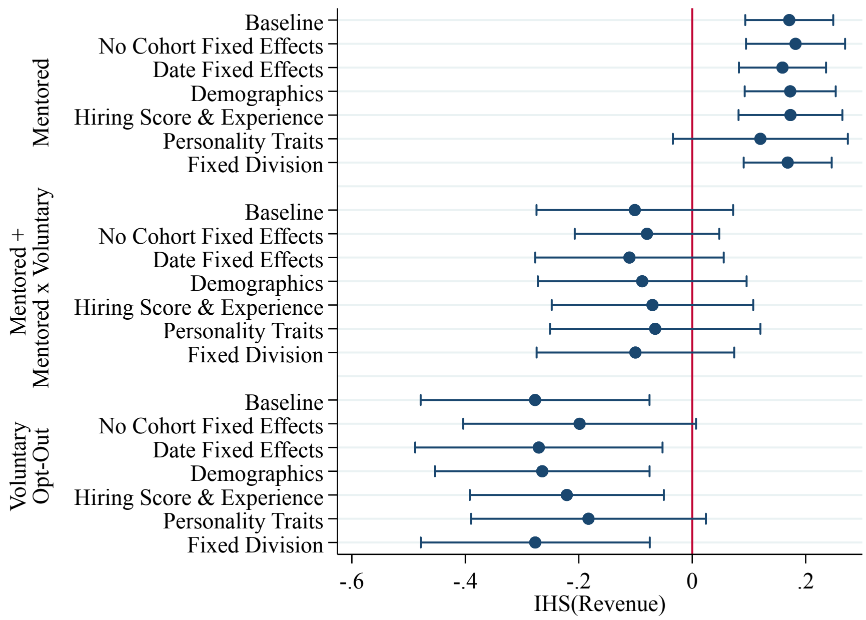
[4] These estimations do not include cohort fixed effects due to limitations in the implementation of the *leebounds* command, reducing their comparability to our main results.

I.A.5 Belief- and Preference-Based Explanations for Opt-Out Behavior

It is possible that beliefs, preferences, or both explain why some agents opt out of the voluntary program. We cannot determine whether miscalibrated beliefs explain the opt-out decision, as we made a design decision to make the program appear as natural as possible. In particular, we did not ask agents in the Voluntary-Condition to report their prior beliefs about program efficacy before soliciting their decision to opt in or opt out due to the concern that reflecting on program benefits may have altered sorting patterns relative to most sign-up procedures that do not ask for such reports. However, our representative survey (see Section 5.2 in the main text) suggests that many workers, approximately 26%, believe similar programs would not benefit them. Accordingly, beliefs around program efficacy may play a role in limiting participation in voluntary training and mentoring programs.

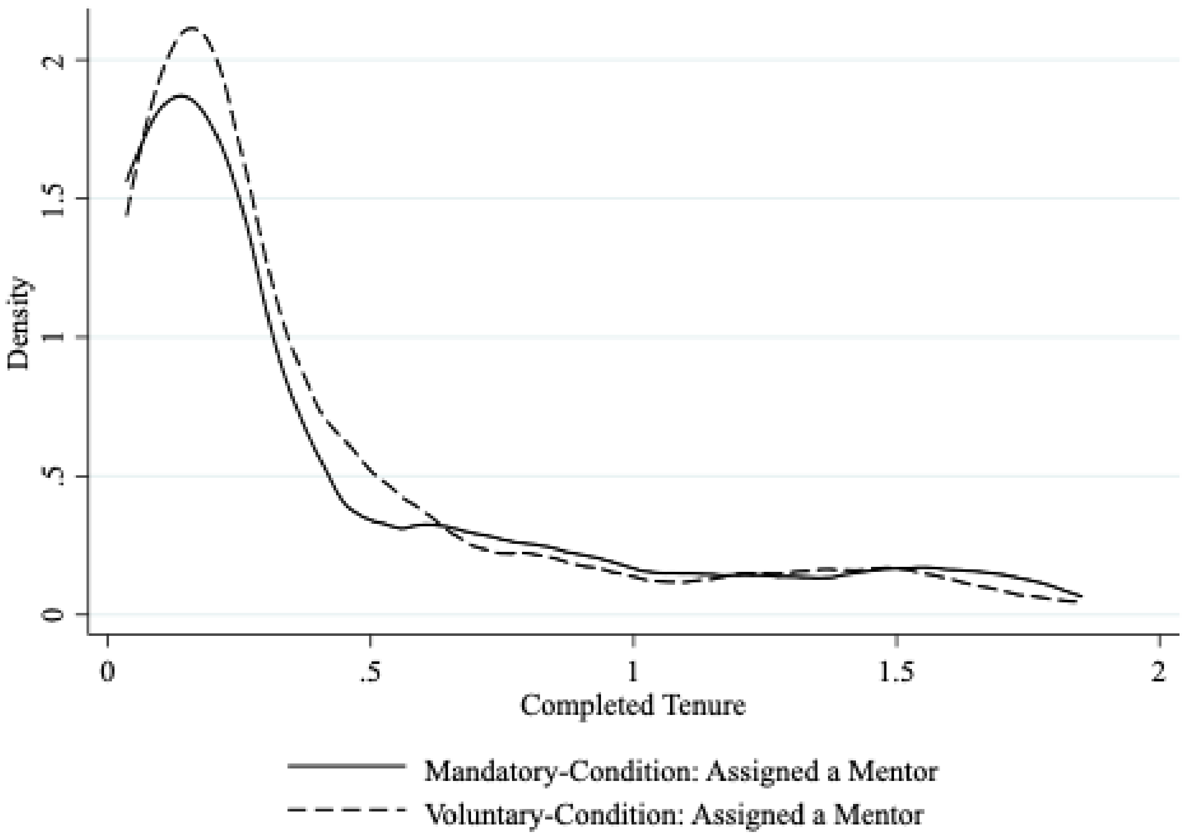
We find limited evidence that workers' preferences explain the decision to opt out, albeit our tests are only indirect. First, to the extent that personality characteristics proxy for preferences across worker types, we show in Table 4 that variation in personality characteristics does little to explain opt-out decisions. Second, we show in Table 6 that most agents in the Mandatory-Condition met with their mentors multiple times, suggesting that an aversion to meeting with more seasoned coworkers is also an unlikely explanation for their opt-out decisions.

Figure I.A.1: Robustness of the Treatment and Selection Effects of Mentoring on Productivity



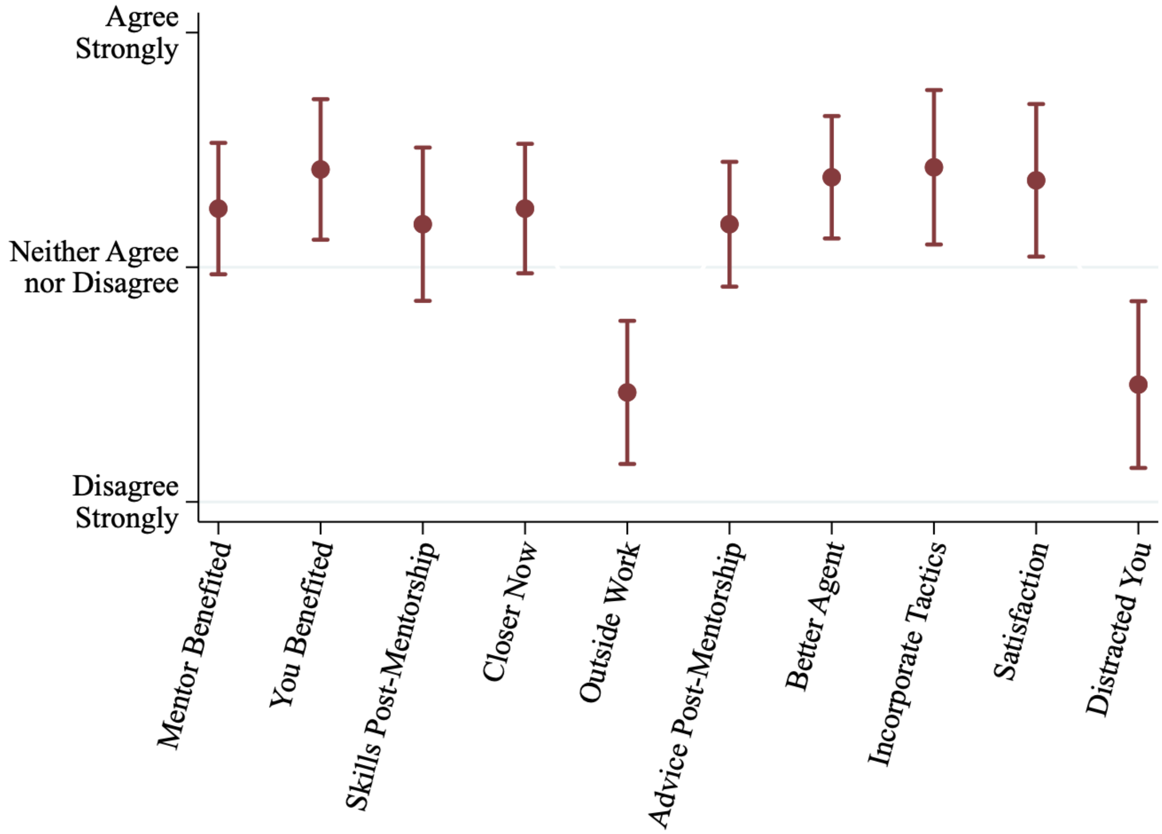
Notes. This figure plots the regression coefficients (and 95% confidence intervals) on *Mentored*, the sum of *Mentored* and *Mentored* $\times$ *Voluntary*, and *Voluntary Opt-Out* from Equation (3). We use IHS(Revenue) as the dependent variable. The "Baseline" estimation replicates the result from Column (7) of Table 2. The second estimation excludes cohort fixed effects. The third includes date fixed effects. The fourth estimation includes controls for the agent's demographic characteristics: age, gender, and marital status. The fifth estimation includes additional controls for the agent's referral status, hiring score, previous call center experience, and previous sales experience. The sixth estimation layers on five more controls for the agent's personality traits: extroversion, agreeableness, conscientiousness, emotional stability, and openness. Note that these controls are frequently missing for non-mentored agents in the Mandatory-Condition, suggesting these are not good controls for estimating the average treatment effect because the data come from a highly selected sample of agents in the Mandatory-Condition who took the intake survey. The seventh estimation removes observations in which agents are no longer working in the division in which they were initially hired.

Figure I.A.2: Tenure of Mentored Agents



Notes. This figure plots the distributions of completed tenure, in years, for each mentored agent in the Mandatory-Condition (solid line) and each mentored agent in the Voluntary-Condition (dashed line). For all agents, the completed tenure is calculated as the difference between their hire date and the date of the last day they are observed in the data, divided by 365.

Figure I.A.3: Responses to Post-Mentorship Survey



Notes. This figure plots the average values (and 95% confidence intervals) for responses to the post-mentorship survey questions. All responses were made on a scale from -3 to 3, with -3 indicating "Disagree Strongly," 0 indicating "Neither Agree nor Disagree," and 3 indicating "Agree Strongly." The statements, from left to right, are as follows: "Your mentor benefited from the mentoring relationship"; "You benefited from the mentoring relationship"; "Since your formal meetings have ended, your mentor has continued to teach you skills to help you make more sales"; "You and your mentor are closer now than you were during the mentor program"; "Since your formal meetings have ended, you and your mentor have spent time together outside of the office"; "Since your formal meetings have ended, you have reached out to your mentor for help/advice"; "You have become a better sales agent as the result of being mentored"; "Being mentored helped you incorporate important selling tactics into your day-to-day work"; "Having a mentor increased your day-to-day satisfaction at work"; "Being mentored distracted you from reaching your potential each week." Seventeen protégés completed the post-mentorship survey.

Table I.A.1: Balance in Mentor Demographics

	Mandatory-Condition	Voluntary-Condition	p-value
	(1)	(2)	(2)−(1)
Mentor Age (yrs.)			
Mean	22.92	23.28	0.732
Std Dev.	(2.57)	(3.20)	
Mentor Female			
Mean	0.07	0.17	0.391
Mentor Married			
Mean	0.14	0.22	0.587
Mentor Tenure			
Mean	1.40	1.12	0.398
Std Dev.	(0.52)	(1.14)	
Number of Protégés	127	155	

Notes. In this table we report average characteristics of the agents who mentored protégés in the Mandatory-Condition in Column (1) and of the agents who mentored protégés in the Voluntary-Condition in Column (2). Mentors were not designated exclusively to either of the mentoring conditions. In other words, a mentor's first protégé could have been assigned to the Voluntary-Condition, whereas their second protégé could have been assigned to the Mandatory-Condition. Mentors were never informed as to whether their protégés were in the Mandatory-Condition or the Voluntary-Condition. We report standard deviations in parentheses, and we report p-values from difference-in-means tests to compare values across the different treatment conditions.

Table I.A.2: Treatment and Selection Effects of Mentoring (Pilot vs. Post-Pilot)

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.180***	0.110***	-0.154	-0.152			0.180***	0.110***	0.174***
	(0.010)	(0.008)	(0.166)	(0.113)			(0.010)	(0.008)	(0.032)
Mentored × Post	-0.033	0.005	0.099	0.096			-0.033	0.005	-0.105
	(0.133)	(0.096)	(0.184)	(0.124)			(0.128)	(0.092)	(0.117)
Voluntary Opt-Out					-0.331***	-0.325***	-0.223**	-0.177**	-0.188***
					(0.065)	(0.009)	(0.092)	(0.085)	(0.023)
Voluntary Opt-Out × Post					-0.063	0.102	-0.225	-0.043	0.028
					(0.186)	(0.103)	(0.228)	(0.153)	(0.069)
Mentored × Voluntary							-0.374**	-0.293**	-0.325***
							(0.143)	(0.106)	(0.026)
Mentored × Voluntary × Post							0.186	0.130	0.245**
							(0.209)	(0.150)	(0.119)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.028	0.032	0.020	0.046	0.047	0.066	0.026	0.040	0.059
Observations	6,725	6,725	7,569	7,569	4,734	4,734	15,670	15,670	15,670

Notes. This table is structured similarly to Table 2, while allowing us to test whether estimates differ between the pilot and post-pilot data as described in Section 3.4. The sample is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity across their first two months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. *Post* equals one for cohorts that entered the firm on or after May 27th (the post-pilot cohorts), and zero otherwise. The specifications in Columns (1) and (2) include all agents in the Mandatory-Condition. The specifications in Columns (3) and (4) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.3: Balance in Division Performance

	Mandatory-Condition	Voluntary-Condition	p-value
	(1)	(2)	(2)−(1)
Revenue			
Mean	803.35	846.61	0.267
Std Dev.	(108.52)	(151.9)	
RPC			
Mean	48.50	51.61	0.305
Std Dev.	(9.31)	(11.39)	
RPH			
Mean	128.45	132.9	0.432
Std Dev.	(16.46)	(21.87)	
Calls			
Mean	17.44	17.41	0.928
Std Dev.	(1.06)	(1.42)	
Hours			
Mean	6.25	6.34	0.638
Std Dev.	(0.68)	(0.66)	
Adherence			
Mean	0.87	0.88	0.272
Std Dev.	(0.02)	(0.03)	
Conversion			
Mean	0.24	0.23	0.285
Std Dev.	(0.03)	(0.02)	
Number of Cohorts	21	31	

Notes. In this table, we take average productivity measures of agents who were not mentorship-eligible within each sales division. Cohorts are assigned to a particular sales division, so these tests estimate the balance in division-level productivity measures between cohorts in the Mandatory-Condition versus those in the Voluntary-Condition. We report standard deviations in parentheses, and we report p-values from difference in means tests to compare values across the different treatment conditions.

Table I.A.4: Treatment and Selection Effects of Mentoring on Productivity (Log-Transformed Dependent Variables)

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.161***	0.101***	-0.080	-0.079			0.161***	0.101***	0.146***
	(0.038)	(0.025)	(0.076)	(0.049)			(0.037)	(0.024)	(0.053)
Voluntary Opt-Out					-0.346***	-0.241***	-0.261***	-0.145***	-0.143***
					(0.099)	(0.063)	(0.089)	(0.046)	(0.028)
Mentored × Voluntary							-0.257***	-0.189***	-0.205***
							(0.082)	(0.056)	(0.054)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.029	0.036	0.021	0.054	0.049	0.072	0.027	0.045	0.059
Observations	6,725	6,725	7,569	7,569	4,734	4,734	15,670	15,670	15,670
p-value: Mentored + Mentored × Voluntary							0.246	0.119	0.151

Notes. The sample used is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity on their first two months on the job after they complete training. Log(.) indicates a variable that is transformed by the logarithm of one plus the value. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to received an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include all agents in the Mandatory-Condition. The specifications in Columns (3) and (4) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of Log(Revenue), Log(RPC), Log(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.5: Tests for SUTVA Violations Comparing New Hire Productivity Relative to Veteran Employees

Panel A: Discouragement/Leakage Tests Between Mentor-Eligible and Hold-Out Cohorts

	IHS(Revenue)		IHS(RPC)	
	(1)	(2)	(3)	(4)
New Hire	-0.354***	-0.275**	-0.431***	-0.393***
	(0.113)	(0.119)	(0.070)	(0.072)
New Hire × Mandatory	-0.025	-0.041	-0.033	-0.044
	(0.117)	(0.120)	(0.069)	(0.070)
New Hire × Voluntary	0.007	0.002	0.047	0.039
	(0.109)	(0.111)	(0.068)	(0.068)
New Hire × Mandatory × Mentored	0.191***	0.199***	0.111***	0.115***
	(0.047)	(0.047)	(0.034)	(0.034)
New Hire × Voluntary × Mentored	-0.067	-0.041	-0.067	-0.050
	(0.067)	(0.067)	(0.042)	(0.042)
Division-Location-Date FE	✓	✓	✓	✓
Demographic Controls		✓		✓
Adj. R-Square	0.073	0.076	0.096	0.099
Observations	47,803	47,766	47,803	47,766
New Hire$_{Man}$ $= 0$, New Hire$_{Vol}$ $= 0$	0.954	0.907	0.445	0.425
New Hire$_{Vol}$ $= 0$, Mentored$_{Vol}$ $= 0$	0.564	0.800	0.281	0.500

Panel B: Crowdout Tests Between Mentor-Eligible and Pre-Experimental Cohorts

	IHS(Revenue)		IHS(RPC)	
	(1)	(2)	(3)	(4)
New Hire	-0.400***	-0.370***	-0.399***	-0.394***
	(0.088)	(0.091)	(0.057)	(0.059)
New Hire × Mandatory	0.009	0.008	-0.074	-0.070
	(0.113)	(0.115)	(0.069)	(0.070)
New Hire × Voluntary	0.027	0.026	-0.015	-0.012
	(0.116)	(0.116)	(0.075)	(0.074)
New Hire × Mandatory × Mentored	0.166***	0.170***	0.093**	0.094**
	(0.062)	(0.065)	(0.039)	(0.040)
New Hire × Voluntary × Mentored	-0.070	-0.057	-0.069	-0.060
	(0.067)	(0.068)	(0.044)	(0.045)
Division-Location-Date FE	✓	✓	✓	✓
Demographic Controls		✓		✓
Adj. R-Square	0.102	0.103	0.131	0.132
Observations	75,094	75,094	75,094	75,094
New Hire$_{Man}$ $= 0$, New Hire$_{Vol}$ $= 0$	0.971	0.971	0.475	0.513
New Hire$_{Vol}$ $= 0$, Mentored$_{Vol}$ $= 0$	0.543	0.695	0.162	0.275

Notes. This table reports tests of the net effect of discouragement, leakage, and crowd-out by comparing the performance of three groups of agents to seasoned veterans who began working at the firm prior to the onset of the experiment: (1) new hires who were in mentor-eligible hiring cohorts; (2) new hires who were in hold-out cohorts that were not eligible for mentorship during the time of the experiment (Panel A); and (3) new hires who entered the firm before the experiment (Panel B). We test whether the performance of treatment-eligible new hires differs from those of non-treatment eligible new hires. Agents who opt out in the Voluntary-Condition are included (see the text for parameter interpretation given this sample). The dependent variable is IHS(Revenue) in Columns (1)–(2) and IHS(RPC) in Column (3)–(4). In the bottom two rows, $NewHire_{Man}$ stands for new hire in the Mandatory-Condition, $NewHire_{Vol}$ stands for new hire in the Voluntary-Condition, and $Mentored_{Vol}$ stands for new hire who was mentored in the Voluntary-Condition. All specifications include division-by-location-by-date fixed effects. Columns (2) and (4) control for agent age, gender, and marital status. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.6: Long-Term Treatment and Selection Effects of Mentoring on Productivity

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.079	0.051	-0.016	-0.053			0.079	0.051	0.055
standard errors	(0.083)	(0.066)	(0.111)	(0.073)			(0.080)	(0.063)	(0.048)
sharpened q-value							[0.329]	[0.372]	[0.315]
Voluntary Opt-Out					0.127	0.075	0.037	0.051	-0.057
standard errors					(0.161)	(0.088)	(0.102)	(0.066)	(0.093)
sharpened q-value							[0.561]	[0.372]	[0.408]
Mentored × Voluntary							-0.099	-0.108	-0.069
standard errors							(0.149)	(0.104)	(0.074)
sharpened q-value							[0.404]	[0.329]	[0.329]
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.040	0.057	0.046	0.035	0.042	0.053	0.043	0.051	0.049
Observations	5,815	5,815	6,608	6,608	3,874	3,874	13,238	13,238	13,238
p-value: Mentored + Mentored × Voluntary							0.852	0.432	0.803

Notes. The sample used is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity across their third to sixth months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include all agents in the Mandatory-Condition. The specifications in Columns (3) and (4) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. Sharpened q-values that adjust for the false discovery rate are presented in brackets, following Anderson (2008). The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.7: Longer-Term Treatment Effects on Retention

	Both Conditions (All Agents)			
	Tenure_{90}	Tenure_{120}	Tenure_{150}	Tenure_{180}
	(1)	(2)	(3)	(4)
Mentored	-0.061	-0.028	0.016	0.005
	(0.076)	(0.095)	(0.069)	(0.088)
Voluntary Opt-Out	-0.242***	-0.213**	-0.157**	-0.094
	(0.075)	(0.080)	(0.065)	(0.062)
Mentored × Voluntary	0.003	-0.010	-0.047	-0.026
	(0.093)	(0.116)	(0.087)	(0.091)
Cohort Fixed Effects	✓	✓	✓	✓
Adj. R-Square	-0.003	0.005	0.003	0.013
Observations	591	591	591	591
Mean Value of Tenure_i	0.45	0.35	0.30	0.26
p-value: Mentored + Mentored × Voluntary	0.288	0.571	0.551	0.658

Notes. The sample used is composed of a single observation per agent, among all mentor-eligible agents with post-training productivity data. $Tenure_i$ equals one for agents who remain with the firm for at least i days after their hire date, and zero otherwise. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in all columns include agents from both conditions. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. The penultimate row reports the average value of $Tenure_i$ for the sample of agents used in the specification within that column. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.8: Determinants of Program Opt-Out and the Relationship Between Opting Out, Productivity, and Worker Characteristics (using Raw Hiring Scores)

Dep. Variable	= 1 if Opted Out				IHS(Revenue)		IHS(RPC)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.002	0.001	0.001	-0.001	-0.003	-0.005	-0.002
	(0.002)	(0.002)	(0.001)	(0.003)	(0.005)	(0.005)	(0.003)
Female	-0.061**	-0.008	-0.004	-0.016	-0.113	-0.101	-0.092
	(0.029)	(0.029)	(0.032)	(0.031)	(0.134)	(0.123)	(0.081)
Married	-0.014	-0.022	-0.020	0.001	0.077	0.125	0.109
	(0.051)	(0.059)	(0.056)	(0.050)	(0.156)	(0.185)	(0.124)
Hiring Score	-0.625**	-0.534**	-0.529**	-0.537***	2.571**	2.766***	2.085***
	(0.255)	(0.212)	(0.206)	(0.204)	(0.969)	(0.806)	(0.466)
Location 1	-0.059	0.040	0.033	0.021			
	(0.081)	(0.053)	(0.052)	(0.052)			
Referral	-0.023	0.019	0.015	0.010	-0.199	-0.189	-0.084
	(0.037)	(0.048)	(0.050)	(0.043)	(0.121)	(0.112)	(0.062)
Call Center Exp.		0.093	0.092*	0.098*		0.400**	0.318**
		(0.062)	(0.054)	(0.054)		(0.187)	(0.113)
Sales Experience		0.000	0.000	0.001		0.005	-0.038
		(0.058)	(0.055)	(0.056)		(0.217)	(0.119)
High Extroversion			-0.030	-0.031		0.206	0.161
			(0.041)	(0.042)		(0.126)	(0.095)
High Agreeableness			-0.040	-0.041		-0.196*	-0.103
			(0.026)	(0.026)		(0.103)	(0.085)
High Conscientiousness			-0.056	-0.054		-0.045	-0.057
			(0.049)	(0.048)		(0.112)	(0.077)
High Emotional Stability			0.042	0.041		-0.151	-0.063
			(0.050)	(0.051)		(0.101)	(0.056)
High Openness			0.023	0.026		0.066	0.082
			(0.039)	(0.038)		(0.132)	(0.086)
Missing Survey				0.285***		-0.439**	-0.288**
				(0.051)		(0.180)	(0.121)
Division Fixed Effects		✓	✓	✓			
Cohort Fixed Effects					✓	✓	✓
(Pse.) R-Square	0.035	0.063	0.078	0.204	0.061	0.074	0.096
Observations	365	341	341	365	4,734	4,734	4,734

Notes. The sample in Columns (1)–(4) is restricted to the 365 agents in the Voluntary-Condition, including those who quit before they completed training. The dependent variable is an indicator that equals one if the agent opted out. The coefficients capture the marginal effects from logistic regressions of different potential predictors of the choice to opt out. Experience and personality factors were collected via survey. We split personality scores on the sample median. Column (4) includes agents who did not complete the new hire survey, which we account for with a *Missing Survey* indicator. In Columns (5)–(6), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(Revenue), the inverse hyperbolic sine of daily revenue, on agents' characteristics. In Column (7), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(RPC), the inverse hyperbolic sine of daily revenue-per-call, on agents' characteristics. In Columns (2)–(4), the Division Fixed Effects indicators reflect the inclusion of controls for whether the agent works in the first largest division, the second largest division, or one of the other smaller divisions. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level). We report marginal effects and delta-method standard errors in parentheses in Columns (1)–(4). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.9: Differences in Worksheet Content Across Conditions in the Experiment

	Total Words per Worksheet	Skill Words per Worksheet	Support Words per Worksheet	Other Words per Worksheet
	(1)	(2)	(3)	(4)
Mandatory-Condition	-0.880	0.357	0.126*	-1.362
	(2.506)	(0.286)	(0.076)	(2.412)
Adj. R-Square	-0.005	0.003	0.010	-0.004
Observations	172	172	172	172
Mean DV	46.94	4.22	0.43	42.29

Notes. This table considers differences in worksheet content between protégés in the Mandatory-Condition and those in the Voluntary-Condition. We have completed worksheet data for 172 mentored agents, as some worksheets that were turned in to the internal mentoring staff were never returned to us. For each agent, we compute the number of words written on all of their completed worksheets, and we divide this by the number of worksheets received from that agent. For each worksheet, we identify the number of words in the responses that relate to job-specific skills or knowledge (*Skill*), those that relate to receiving support, encouragement, and friendship (*Support*), and those that are neither related to skill nor support (*Other*), which include stop words. These become the dependent variables in our regression specifications of worksheet content on mentorship type. Robust standard errors are reported in parentheses. The mean of the dependent variable is listed below the observation count line. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.10: Productivity Treatment Effects When a Mentorship Slot is the Unit of Analysis

Panel A: Months 1–2

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.141***	0.108***	-0.069	-0.050			0.141***	0.108***	0.114***
	(0.053)	(0.033)	(0.053)	(0.036)			(0.053)	(0.033)	(0.018)
Voluntary Opt-Out					-0.276***	-0.162***	-0.199***	-0.097**	-0.077***
					(0.064)	(0.047)	(0.061)	(0.043)	(0.027)
Mentored × Voluntary							-0.232***	-0.171***	-0.149***
							(0.084)	(0.053)	(0.029)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.023	0.034	0.025	0.043	0.036	0.046	0.024	0.036	0.078
Observations	10,216	10,216	9,986	9,986	6,588	6,588	22,451	22,451	22,451
p-value: Mentored + Mentored × Voluntary							0.346	0.345	0.433

Panel B: Months 1–6

	Mandatory-Condition (All Agents)		Voluntary-Condition (Opt-In Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	0.056*	0.043**	-0.028	-0.023			0.056*	0.043**	0.054***
	(0.031)	(0.021)	(0.035)	(0.024)			(0.031)	(0.021)	(0.019)
Voluntary Opt-Out					-0.116***	-0.070**	-0.094**	-0.035	-0.036*
					(0.038)	(0.027)	(0.036)	(0.026)	(0.021)
Mentored × Voluntary							-0.093*	-0.074**	-0.071***
							(0.050)	(0.034)	(0.026)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.013	0.021	0.019	0.026	0.029	0.036	0.018	0.026	0.078
Observations	27,920	27,920	28,569	28,569	19,070	19,070	63,242	63,242	63,242
p-value: Mentored + Mentored × Voluntary							0.432	0.351	0.426

Notes. The results in this table show estimates of the treatment effects of mentorship when a slot (i.e., an occupied position in either the treatment or control group during the experiment) is the unit of analysis. This enables us to track long-term performance and account for attrition. To do so, we form a panel that is "balanced" by taking the actual productivity of agents when they remain at the firm, while replacing the productivity measure with the expected productivity of imputed replacement agents if the agents have left. In other words, for mentor-eligible agents who leave the firm before the two-month mark (six-month mark) in Panel A (Panel B), we extend the time series of their productivity provision to two (six) months and replace their post-termination productivity values with the productivity of a randomly newly hired replacement agent. Specifically, use the productivity of a randomly chosen newly hired non-mentor-eligible agent in the same location-division-year-quarter as the departed agent. We then re-estimate our main intention-to-treat regression models. We estimate ordinary least squares regressions with cohort fixed effects in all columns. We conduct 200 iterations of this estimation procedure to bootstrap the standard errors. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.11: Lee Bounds Estimates of the Effect of Mentoring on Productivity

	Mandatory-Condition		Voluntary-Condition	
	IHS(Revenue)	IHS(RPC)	IHS(Revenue)	IHS(RPC)
	(1)	(2)	(3)	(4)
Mentored_{lower}	0.169***	0.094***	-0.222***	-0.131***
	(0.046)	(0.035)	(0.076)	(0.048)
Mentored_{upper}	0.208**	0.122**	-0.058	-0.018
	(0.083)	(0.051)	(0.042)	(0.031)
Observations	8,408	8,408	9,231	9,231

Notes. This table uses agent-day productivity data for agents in the Mandatory-Condition in Columns (1) and (2) and for agents in the Voluntary-Condition, excluding those agents who opt out of the program, in Columns (3) and (4). The sample is a balanced panel of agents' productivity through their first two months of tenure, augmented with missing data for those who exit the firm during the panel. We estimate treatment effect bounds that account for non-random attrition as proposed by Lee (2009). These estimations do not include cohort fixed effects, as the *leebounds* command cannot accommodate fixed effects, which causes the estimates to differ from those in our main results. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table I.A.12: Tests for Variation in Mentoring Treatment Effects Based on Exposure to Teammates with the Same or Different Treatment Status

	IHS(Revenue)	IHS(RPC)
	(1)	(2)
Mentored	0.154***	0.101***
	(0.035)	(0.019)
Mentored × Same Treatment	0.030	0.022
	(0.176)	(0.076)
Cohort FE	✓	✓
Adj. R-Square	0.033	0.036
Observations	6,725	6,725

Notes. The sample consists of agents in the Mandatory-Condition. After training, agents joined different teams, after which they would begin the mentoring program. For each agent, mentored or not, we generate an intermediate variable (not the regressor), *Same Teammate*, which equals one if, on that day, all the agent's teammates from their hiring cohort had the same mentored/non-mentored treatment designation as their own. Otherwise, *Same Teammate* equals 0 for that agent-day. We then compute the rolling average of the *Same Teammate* variable for each agent, from their first day on the job after training to the present day, and we label this *Same Treatment*. The variable *Same Treatment* captures the fraction of days on the job historically that a new hire worked alongside teammates from their same hiring cohort who had the exact same mentored/non-mentored treatment designation as themselves. So, a mentored (non-mentored) agent with *Same Treatment* = 1 has never worked alongside a non-mentored (mentored) teammate from their same hiring cohort and would have no perceptions of differential treatment. The mean of *Same Treatment* in the sample is 0.22. All specifications include cohort fixed effects. Standard errors are clustered by cohort for those workers entering the experiment after the pilot-period and by pilot-period-by-office for those workers entering during the pilot (this is because the pilot entailed assignment of the Mandatory- and Voluntary-Conditions at the office level), and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

I.B AEA Pre-Registration Text

Here we replicate the AEA pre-registration text. Differences between the AEA pre-registration and our actual implementation are denoted in footnotes.[5]

I.B.1 Abstract

Mentoring is increasingly encouraged in workplaces, and a number of firms have implemented formal programs. While a growing body of research suggests that mentoring relationships benefit those being mentored (protégés), there is scant evidence to delineate whether these favorable outcomes are driven by the mentoring experience on average, by the self-selection of protégés into mentoring who anticipate having the largest gains (selection based on gains), or by the self-selection of protégés who would have performed well in the absence of mentoring (selection based on levels). We use a field experiment to evaluate a workplace mentoring program inside a large sales organization.

Experienced employees opt-in as mentors, and new hires are slated as potential protégés. The project objective is to study the mentoring consequences across protégés who actively elect to be formally mentored relative to those who are randomly allocated a mentor. We estimate treatment effects on sales productivity and turnover for those who select into mentoring and for those who opt out.

I.B.2 Intervention(s)

We analyze the effectiveness of a workplace mentoring program where employees opt-into mentoring or are randomly assigned a mentor. More details are provided in the design field.

I.B.3 Intervention Start Date

2019-05-27

I.B.4 Intervention End Date

2019-12-20

I.B.5 Primary Outcomes (end points)

Log revenue-per-call (RPC), an indicator for worker turnover, log completed tenure, the firm's internal adherence to schedule measure (e.g. time spent working whilst at work), and the firm's internal engagement metrics (online surveys asking for willingness to recommend employment at firm, comfort with leadership, etc.).[6]

I.B.6 Primary Outcomes (explanation)

Agent's weekly RPC is a measure of sales productivity that removes demand variation outside of the worker's control. RPC is the primary productivity measure used by the firm, combining

[5] We received IRB approval at all of our respective institutions prior to the onset of the mentoring program. See IRB19-0769.

[6] Following our prior paper, we use an approximate logarithmic transformation of RPC (and revenue). In this case, we use IHS() rather than logs because we have daily data that includes cases where these newly hired agents make zero sales. Our results, as shown in Table I.A.4, are similar if we use the original pre-registered log() measure.

both agent's firm-specific knowledge and their individual effort.[7] Worker turnover measures whether the interventions changed the agents' propensity to leave the firm. Log of completed tenure is a different measure of retention that has been used in the prior literature and the attendance measure provides an adjacent measure of agent effort.[8] Finally, engagement measures are hypothesized to be forward looking measures of productivity.[9]

I.B.7 Experimental Design Details

Seasoned sales agents are invited to apply as internal mentors to incoming recruits (the firm "qualifies" mentors as having sufficient sales experience). New mentorship opportunities are periodically announced, and prior mentors are permitted to re-enter the mentor pool. The firm communicates that serving as a mentor is a useful first step to being considered for a managerial position. New mentors complete a survey asking them about their personality, interests, work preferences, and values. Mentors are randomly assigned with probability 50% to receive a set of instructions emphasizing that mentoring is about teaching protégés how to do the job. The remaining mentors receive instructions emphasizing that mentoring is about providing protégés support. Sales agents are hired in batches (cohorts). Newly hired sales agents complete two weeks of training, primarily in a classroom or listening in on other agents' sales calls. New agents then complete the same personality and preference survey that mentors take. At the end of their two-week training, each cohort of agents is eligible for randomization into a mentoring treatment arm. Any mentoring relationship commences as soon as the agent completes their training.

The randomization procedure is as follows:

I.B.8 Cohort Level Randomization

The initial level of randomization is cohorts of new hires (potential protégés). Each cohort (a group of new hires who are joining the firm at the same time, are in the same training group, and will be working in the same sales division and office location) will be randomized into one of two conditions: Mandatory-Condition or Voluntary-Condition. 40% of the cohorts will be in the Mandatory-Condition group and 60% of the cohorts will be in the Voluntary-Condition group.

I.B.9 Within Cohort Randomization

For cohorts in Mandatory-Condition, new hires will receive a mentor with probability 50%. This will be communicated privately between sales floor staff and the individual workers. Agents in the Mandatory-Condition who do not receive a mentor will not receive communication regarding the program. For cohorts in Voluntary-Condition, sales floor staff verbally explain the firm's mentorship program, answer questions, and provide each agent a confidential ballot where they can decide whether or not to enter a lottery which randomly

[7] RPC was the primary endpoint based on our experience analyzing the productivity of veteran agents within the firm (Sandvik et al., 2020), but seminar participants asked us about other margins of adjustment (like time use) that is captured by total revenue and is more relevant for new agents compared to more experienced workers, which is why we report both metrics.

[8] In practice, we find no differences in completed tenure and simplify the analysis by using indicators for completed tenure rather than log completed tenure.

[9] As noted below, the firm changed the cadence of collecting engagement metrics (which were planned for 5 weeks after training completion), so we do not have these measures for many cohorts.

determines whether the agent is allocated a randomly assigned mentor, or no mentor at all. Of the agents who enter the lottery, approximately 50% will be assigned a mentor. Agents who choose not to be mentored will never be assigned a mentor.

I.B.10 Compliance Tracking

The firm's training staff will track whether mentors and protégés meet. This tracking will be aided by worksheets. Upon completion of the worksheets, the firm will reward "kudos" points that can be accumulated to purchase items from the company store. As mentioned earlier, mentors may participate more than once, however they will never have more than one protégé at a time.[10] Eligible protégés and mentors will each take an electronic survey at the end of the formal program. The survey for protégés will ask about the protégé's initial excitement when told about the mentoring program, their perceived engagement with their mentor, and an estimate of the effectiveness of mentoring. This question will be phrased as: "What was your average RPC last week? What do you think your average RPC would have been had you not been working with a mentor?" The survey for mentors will ask about the protégé's enthusiasm for the mentorship program and an estimate of the mentor's perceived treatment effect on the protégé. This question will be phrased as: "If your protégé had not received mentoring, his/her RPC would have been [40% lower — slider — 40% higher].[11] Note that numbers greater than zero mean that mentoring was not effective for improving protégé performance. Please be candid, as your responses will not be shared with management."

I.B.11 Edit June 4, 2019

To assess the potential for spillovers, we have revisited the design in consultation with the company such that there will be "hold out" cohorts for one division-office who never receive mentoring. Any cohorts/individuals who are switching brands also will be held-out. Work-from-Home cohorts will also present a possible "hold out" group for comparison and all cohorts in a smaller third office (which no longer exists, but for whom historical data is available) were "hold out" cohorts who knew nothing about mentoring.[12] A "sentiment survey" will be administered to all agents in their 5th week on the sales floor.[13] This will be one week after mentored agents finish hiring. We will gather information on their feelings towards the onbaording process and ask questions, common in the literature, to solicit their sentiment towards the firm, their perceptions of their ability, their enthusiasm about the job, etc. We will use this survey to test for spillovers based on survey responses.

[10] As the program progressed, the internal mentoring staff felt that many of the mentors could effectively mentor multiple protégés as once. As a result, we adjusted the protocol such that it was possible for a single individual to mentor multiple new hires concurrently, but mentor-protégé pairs always met individually, meaning the protocol was the exact same from the point of view of the protégé.

[11] The post-mentorship survey completion rates of mentors and protégés were very poor, so we do not have meaningful data for this question. Anecdotally, the average responses of both sets of individuals suggests that protégés' RPC would have been lower in the absence of mentorship, but the inference is not precise.

[12] Further consultation with the firm's staff meant that the hold-out procedure was not isolated to a single division or office but instead involved rotation of some cohorts out of the program. More detail about hold-out cohorts is provided in the main text.

[13] We were not able to administer this survey. The firm had several of its own survey initiatives occurring simultaneously, so additional surveys connected to the mentorship program were not conducted due to the concern of "survey fatigue" among the sales agents.

I.B.12 Randomization Method

Randomization done by computer. Participants will be informed if randomized in.

I.B.13 Randomization Unit

Clustered randomization of cohorts in a first level, with individual randomization within the cohort. See design details.

I.B.14 Was the treatment clustered?

Yes

I.B.15 Sample size: planned number of clusters

The exact sample size is stochastic and depends on the firm's actual hiring. We have 46 planned clusters.

I.B.16 Sample size: planned number of observations

In one office, the firm has projected 269 new hires in 22 cohorts. There are 350 new hires in 24 cohorts projected in the second office.

I.B.17 Sample size (or number of clusters) by treatment arms

Please see design field.

I.B.18 Minimum detectable effect size for main outcomes

Using pre-intervention data to estimate the intra-class correlation coefficient and residual variation, the minimum detectable effect size for log RPC between those randomized into and out of mentoring is 0.07 (accounting for sample design and clustering).

I.B.19 Analysis Plan

The Treatment Effect of Mentoring on those who opt in is:

$$\beta_{OptInMentor} = mean(Y_{OptInMentor}) - mean(Y_{OptInNoMentor}).$$

We will estimate this mean difference using a regression of Y on an indicator for receiving a mentor along with cohort fixed effects and indicators for the type of instructions mentors receive.[14] The sample will be the workers in the voluntary treatment cohorts who opt into mentoring.

The Treatment Effect of Mentoring on those who opt out can then be derived by writing the average gain from mentoring in the population as:

$$mean(Y_{RandomMentor}) - mean(Y_{NoMentor}) = \beta_{OptInMentor}\pi_{OptIn} + \beta_{OptOutMentor}\pi_{OptOut}.$$

The β parameters are the heterogeneous treatment effects and the π are the population fraction who opt in and opt out. This yields:

[14] Mentors received instructions that either put more emphasis on the supportive nature of the program or the skills-building nature of the program. We detect no differences between instruction type. Because of this and for brevity, we omit this indicator from the models in our treatment effects tests.

$$\beta_{OptOutMentor} = [mean(Y_{RandomMentor}) - mean(Y_{NoMentor}) - \beta_{OptInMentor}\pi_{OptIn}]/\pi_{OptOut},$$

where the difference in means is net of cohort fixed effects and indicators for mentoring instruction type. The population average treatment effect (ATE) of mentoring can be estimated from a regression of Y on a dummy for receiving a mentor and cohort fixed effects in cohorts that have (entirely) randomly assigned mentoring. This yields:

$$\beta_{OptOutMentor} = [ATE - \beta_{OptInMentor}\pi_{OptIn}]/\pi_{OptOut}.$$

Inference for $\beta_{OptOutMentor}$ will come from block bootstrapping the statistic. Selection bias will be measured among voluntary treatment cohorts as the regression analogue of:

$$mean(Y_{OptInNoMentor} - mean(Y_{OptOut}),$$

where the means are net of cohort fixed effects. This procedure allows us to estimate sales productivity differences among protégés who opt into mentoring and those who do not. We use the sample of agents in the voluntary cohorts who did not receive a mentor. We regress Y on an indicator that the agent opted into mentoring along with cohort fixed effects and their mentor instruction-type fixed effects. Other regressions will look at opt-in as a function of early sales and demographic characteristics (gender, age, office location) and past experience (prior sales or call center experience).

We plan to validate these estimates using the electronic survey responses collected after the protégé graduates from the formal mentoring program, approximately 4 weeks following the initial onboarding instruction (e.g. how to use the systems, enroll for benefits, etc.).[15] We will compare average perceived gains from mentors and protégés to the actual estimated treatment effects across different assignment conditions. We will then assess whether the effectiveness of the mentoring pair differs based on characteristics of the mentor and protégé. We will regress protégé sales on fully saturated interactions of demographic characteristics for the mentor-protégé pair (old/young based on coarse buckets; gender) as well as similarity in survey responses on the intake survey.[16]

Finally, to assess whether mentoring detracts from—or improves sales—for the mentor, we will regress mentor log RPC and other sales measures on indicators demarking whether the mentor is eligible to mentor but has not yet done so, whether they have previously mentored in the program, or whether they are actively mentoring a protégé. This regression will include mentor fixed effects and mentor tenure.[17]

[15] As mentioned earlier, we were not able to administer this survey.

[16] This is a very high-dimensional exercise, and the most interesting potential differences (like females mentoring females) ended up having very small cell sizes, reducing statistical power for these tests.

[17] Tests that compare the characteristics of mentors and protégés, and those that look at the impact of mentorship on mentor productivity, are likely to be discussed in a separate article.

I.C Documentation of Instructions to Mentors and Example Worksheet for Structuring Conversations

Mentor Instructions

What is a Mentor?

In *The Odyssey*, Odysseus prepared to fight in the Trojan War. Before leaving home to fight in the war, he asked his trustworthy friend, named Mentor, to train and educate his son, Telemachus. Similarly, mentors today are meant to train and educate their protégés. Management at has chosen you to be a mentor---a source of further skill development---for newly hired sales agents. You have been selected specifically because you've demonstrated a willingness to teach other sales agents and help them become a successful and productive sales agent.

The responsibility to mentor a newly hired sales agent should not be taken lightly. Management strongly believes new agents will benefit from the additional training and the insider knowledge received as a result of being mentored by a talented, more seasoned agent. Because of this, has devoted significant resources to give mentors and protégés the best opportunity to spend productive time together, so please take your mentorship responsibilities seriously.

What will You Do as a Mentor?

As a mentor, you will do the following:

1. You will meet with your protégé at **least once a week**.
    a. Before meeting, your protégé will complete the Protégé Worksheet.
        i. If he/she has not completed it, you will kindly help him/her do so.
    b. During your meeting, you and your protégé will discuss his/her responses. You should also take this time to do the following:
        i. Impart knowledge and skill by explaining, giving useful examples, and demonstrating processes, and asking thought-provoking questions.
        ii. Discuss actions you've taken to become a successful sales agent.
        iii. Provide him/her with any tips and sales tactics that help you overcome customer concerns and that help you up-sell to better services.
        iv. Practice the designated sales protocol with them and help them gain a strong understand of the products, services, and bundles available.

2. After meeting with your protégé, you will deliver the finished worksheet to .
    a. will initial and timestamp the worksheet and make a record that you completed your weekly meeting responsibility.

3. Every two weeks, you will be asked to complete an on-line survey.
    a. These questions are meant to gauge the progress of your protégé and the overall benefit of the mentoring relationship.
    b. Please answer these questions honestly, as they are not meant to punish but, instead, to help assess the effectiveness of the mentorship program.

Protégé Worksheet (Week 1)

Protégé: ____
Mentor: ________________________________ Number of times mentor has reached out: _____
Date: ____

Weekly Self-Reflection:

What are **your expectations** regarding your sales ability? Does your mentor know this?

What may **prevent you from** having a successful first week? Does your mentor know this?

Think of the **MOST** successful call you had recently. What made it **successful**?

Think of the **LEAST** successful call you had recently. What made it **unsuccessful**?

Weekly Goal:
What **ONE goal** are you setting for yourself for this coming week?

What will you do to **reach this goal**? Have you told your mentor about this goal? __________
____

For Mentors to Respond:
How have you, as a mentor, been a source of **skill development** for your protégé? What have you done so far to help him/her succeed on the sales floor here at ?

Protégé's Initials   Mentor's Initials   Intern's Initials & Timestamp

I.D Pre-Registered Estimation of Heterogeneous Treatment Effects

We can use the estimated Mandatory-Condition and Voluntary-Condition productivity treatment effects, along with the data on the fraction of Voluntary-Condition agents who opt out of receiving a mentor, to estimate the treatment effect of mentorship among opt-out agents. We pre-registered the following procedure for this purpose. Using productivity measure Y, we define the conditional average treatment effect of mentoring given selection into participation as the difference in expected production between mentored and non-mentored agents conditional on opting in:

$$ATE|OptIn = \beta_{OptInMentored} = E(Y_{OptInMentored}) - E(Y_{OptIn\sim Mentored}).$$

We can then express the unconditional average treatment effect of mentorship as the weighted average of heterogeneous effects with shares π:

$$\begin{aligned} ATE = E(Y_{MandatoryMentored}) - E(Y_{Mandatory\sim Mentored}) = \\ \beta_{OptInMentored} \times \pi_{OptIn} + \beta_{OptOutMentored} \times \pi_{OptOut}. \end{aligned}$$

Rearranging terms, we get,

$$\beta_{OptOutMentored} = \{ATE - \beta_{OptInMentored} \times \pi_{OptIn}\}/\pi_{OptOut}.$$

We use the estimated treatment effect in the Mandatory-Condition as the estimated ATE, and we use the estimated treatment effect in the Voluntary-Condition as the estimated $ATE|OptIn$.[18] The values of π come from the proportion of agents who opted out in the Voluntary-Condition. We show the estimated treatment effect for opt-out agents in Table I.D.1, where standard errors come from 500 block-bootstrap iterations by cohort. The point estimate of 1.207 in Column (1) of Panel A implies that opt-out agents would have more than doubled their overall revenue generation, on average, had they received mentorship. Based on this analysis, opt-out agents were those who would have benefited the most from receiving mentorship. That is, program participation is negatively correlated with treatment gains.

[18] We include cohort fixed effects in estimating these treatment effects.

Table I.D.1: Estimated Treatment Effect of Mentoring Among Opt-Out Agents

Pre-Registered Estimates of Opt-Out Treatment Effects		
	IHS(Revenue)	IHS(RPC)
	(1)	(2)
Opt-Out Mentored Effect	1.422**	1.159***
	(0.648)	(0.417)

Notes. This table reports estimates of the treatment effect of mentorship among agents who would have opted out of the program using the pre-registered estimator described above. This estimator imposes that the full difference between the Mandatory- and Voluntary-Condition comes from heterogeneous treatment effects and selection. To estimate standard errors, we block-bootstrap by cohort (N = 52) over the whole procedure, with 500 bootstrap replications for each column. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

I.E Worksheet Response Examples

Panel A: Think of the most successful call you had recently. What made it successful?	
Skill	· I pitched TV really well · Having different examples of pitches from my coach to fall back on
Support	· I was confident and tried to connect · The person I spoke with was very nice

Panel B: Think of the least successful call you had recently. What made it unsuccessful?	
Skill	· Customer didn't want to pay the deposit, [I] didn't rebuttal · Not doing call flow, not caring, not enough discover
Support	· Not being confident in my ability to rebuttal · The person was rude and wanted me fired

Panel C: What will you do to reach this goal? Have you told your mentor about this goal?	
Skill	· I will create better pitches · Be better with the triple play, use what [the] mentor told [me] · My mentor is going to help me pitch DTV by giving me her tips on what helped her · Practice on every unserviceable call · Try upsell technique
Support	· [Goal to achieve] 1500 a day, build confidence in it · Be more positive · Stay positive · Stay in communication with [my coach] · Check in with my coach and be confident

Panel D: Words Associated with Sales Skills and Knowledge
Adherence, Conversion, Customer, Direct, Dish, Double, DPI, DTV, Internet, Knowledge, Phone, Pitch, Price, Pricing, Process, Revenue, RPC, RPH, Sale, Security, Sell, Skill, Sold, System, Television, Triple, TV

Panel E: Words Associated with Receiving Support
Annoy, Breath, Confidence, Confident, Cool, Encourage, Encouraging, Friend, Introduce, Kind, Laugh, Mean, Motivate, Motivation, Nice, Patience, Patient, Positive, Rude, Social, Support, Welcome, Welcoming

I.F Calculation Details for the Value to the Firm

This section details how we calculate the NPV to the firm from the mentorship program. In all calculations, we compare the actual stream of revenues generated by treated employees net of commission pay to the actual stream of revenues generated by control employees, again net of commission pay. To account for potential retention differences and to get at the long-term effect, we assume that each treatment is a discrete investment that happens only once. We then use the treatment or control worker's slot as the unit of analysis, where we track each worker over a six-month horizon. If a worker leaves, we continue to track that worker's shifts going forward, but we fill in their productivity with the expected net revenue of a replacement employee.

To estimate additional revenues, using a treatment event as the unit of analysis, we take the estimated treatment effect in the Mandatory-Condition across the first six months of tenure, $5.75\% = e^{0.056} - 1$ (from Column (1) in Table I.A.10, Panel B), and multiply it by the average daily revenue of agents in non-mentored slots, \$735, yielding a daily revenue increase of about \$42 over the first six months of tenure. There were 127 mentored agents in the Mandatory-Condition, resulting in \$882 more in revenue per agent-month based on agents' total days of work. The firm earns this additional revenue net of an 8% commission rate paid to sales agents. Hourly pay is invariant to productivity and did not change with the program. We multiply these monthly net-revenue amounts by six, the number of months, and by 127, the number of mentorship slots. We conservatively assume this additional revenue is realized at the end of the year and discount the future cash flow using a 12.5% discount rate, which gives us a present value of the additional revenue earned by mentored agents equal to approximately \$550,000.

Second, we consider the costs. These costs include the estimated time costs of taking the mentors off the phone (protégés' opportunity costs are included in the revenue treatment effects) and administrative costs. Mentors and protégés spent 30 minutes in the mentorship meetings each week. Revenue per hour for mentors averaged \$148 and they were paid an additional \$10 of "kudos" points for completing each meeting. Together this implies a cost of \$84 per meeting or \$30,000 total, assuming that each treated agent met with their mentor 2.5 times. We include the administrative costs of the two internal mentorship staff members who oversaw the program in the two locations, estimated to total approximately \$67,000, (generously) assuming that mentoring administration accounted for 50% of their workload. This leads to total costs of about \$97,000. Together, this leads to a net present value of the program equal to approximately \$453,000, which may be a lower bound if more productive agents allow sales managers to have larger spans of control (Espinosa and Stanton, 2021).

The estimate of potential gains from reallocating agents in the Voluntary-Condition to the Mandatory-Condition is done as follows: There are 155 agents to reallocate. Assuming the full treatment effects translate, we can use the previous estimates, which allocate overhead and opportunity costs, and scale by the relative number of agents to get roughly $\frac{155}{127} \times \$453,000 = \$553,000$. We then account for the fact that overhead was already allocated to the Mandatory-Condition, yielding a gain of \$620,000. This number can be scaled by the treatment effect difference due to self-selection and treatment effect heterogeneity.

Finally, the switch from a voluntary to a mandatory program would have led to an increase in gross commissions to agents of around \$65,620 ($= \$42 \times 21 \times 6 \times 0.08 \times 155$).