The Effects of Digital Surveillance and Managerial Clarity on Performance*

Namrata Kala† Elizabeth Lyons‡

January, 2025

Abstract

How managers frame the adoption of organizational practices may impact the returns to such practices, but managerial justification is often correlated with the use of particular practices or other dimensions of managerial quality. Using a randomized control trial, we study how the causal impacts of a frequently used monitoring management practice for remote work employers–digital worker surveillance–varies by randomly allocated justification for its use. In an online labor market, we divide workers into low and high-productivity performers and randomize both whether surveillance is used and whether its use is justified based on the workers' baseline productivity. We find that digital surveillance does not have significant effects on worker performance on average, but that not explaining the presence or elimination of digital surveillance based on worker performance significantly reduces worker output. Our results demonstrate a nuanced relationship between monitoring and worker performance that depends on how monitoring is rationalized to workers.

JEL Codes: J24, J28, J53, M54

*We gratefully acknowledge funding support from J-PAL's Jobs and Opportunity Initiative. We would like to thank Zixu Chen, Pryce Davies, and Isoke Nelson for excellent research assistance. The experiment described and analyzed in this paper has been registered at AEA RCT Registry under AEARCTR-0011365. All errors are our own.

†MIT Sloan School of Management, CEPR, NBER. email: kala@mit.edu

‡School of Global Policy & Strategy, UC San Diego. email: lizlyons@ucsd.edu.

1 Introduction

How managers frame the adoption of technologies may moderate their effects on performance. In particular, to the extent that trust between workers and managers is important for firm performance, providing workers with justification for potentially discouraging managerial decisions such as pay inequality can mitigate the morale effects of these practices (Breza et al., 2018). One commonly used and fast-increasing managerial practice with ex-ante ambiguous returns is the use of worker surveillance technologies. Intensive input-based monitoring could mitigate shirking and improve productivity (Pierce et al., 2015; Kelley et al., 2021), but workers may interpret such surveillance as restricting their autonomy, which could reduce worker engagement in the job and performance as a result (Gambardella et al., 2020). However, if changes in monitoring are understood by workers to be justified rather than arbitrary or uninformed, any negative effects of monitoring on employee motivation and productivity may be mitigated.

The use of worker surveillance technologies was on the rise throughout the 2010s and saw a spike in use during the shift to remote and hybrid work during 2020.[1] Employer surveys indicate that the majority of employers with remote or hybrid employees use digital surveillance tools.[2] These technologies can lower the incidence of moral hazard in tasks for which worker effort is hard to observe via outputs and otherwise hard to observe in the production process (Jensen et al., 2020; Kelley et al., 2021; Staats et al., 2017).[3][4] Theoretically, any performance gains caused by digital surveillance may come at the expense of eroded employee satisfaction (Chang and Lai, 1999; Frey, 1993). Consistent with this, Ko and Baek (2024) find that input-based digital monitoring is associated with a significant improvement in the performance of remote USPTO examiners, but that it is also associated with lower worker retention.

[1] See, for instance, https://www.nytimes.com/wirecutter/blog/how-your-boss-can-spy-on-you/.

[2] https://www.expressvpn.com/blog/expressvpn-survey-surveillance-on-the-remote-workforce/, https://digital.com/6-in-10-employers-require-monitoring-software-for-remote-workers/#:~:text=Sixty%20percent%20of%20employers%20with,spending%20their%20time%20(79%25)

[3] By doing so, they may also reduce barriers to qualifications and skills communication between new worker types and employers. However, if workers have a disutility of being monitored, it may also reduce worker motivation and well-being, with potentially negative impacts on productivity, rendering the net effect of such technologies ambiguous. For instance, Agrawal et al. (2016) demonstrates that the use of digital monitoring is associated with higher rates of low-income workers being hired by high-income country employers.

[4] Similarly, Pierce et al. (2015) find that employee theft digital monitoring systems reduce theft and improved worker productivity.

In this paper, we use a randomized control trial (RCT) to study the returns to worker surveillance, as well as how information justifying managers' decisions to implement digital surveillance affects worker performance and job satisfaction. To answer these questions, we conducted a RCT with 434 remote workers hired on UpWork. Workers who participated in our study were asked to transfer data from poor-quality scans of firm financial statements into spreadsheets.[5] All workers in our sample perform their first two hours on the task while being monitored by digital surveillance software. Worker output performed during these first two hours was evaluated, and workers were assigned to our low or high-performance treatment conditions based on their productivity relative to a baseline pool of workers' sample median productivity.

Subsequently, the experiment randomly varied two aspects of the work environment: first, the working conditions (surveillance or not), and second, whether workers were informed that their surveillance condition was based on their job performance or not. Workers in the low-performance subsample were randomly assigned to receive one of the three messages about surveillance for the final 6 hours of the task. One-third of the subsample was told they would need to continue working with surveillance software activated *because* their productivity was low (*surveillance-justified*), another third was told their productivity was low and were required to continue working with the surveillance software activated but their performance was not used to justify the continued use of monitoring (*surveillance-unjustified*), and the final third were told their productivity was low but were given the option to work without being monitored by the surveillance software (*low-performance control*).[6] Analogously, workers in the high-performance subsample were also randomly assigned to receive one of three messages about surveillance for the final 6 hours of the task. One-third of the subsample was told they no longer had to work while being monitored by the surveillance software *because* they had high productivity during the first two-hour work window (*no surveillance-justified*), another third were told they had high productivity during the first two-

[5]Workers were, accurately, informed that image recognition software does not accurately capture and enter the data and, thus, that they should not use it. We suspect that a small number of other workers used automatic data entry software based on their speed of data entry. We do not find any relationship between potential software use and treatment, our results are robust to dropping them from the sample, and we include an indicator for whether or not we suspect a worker of using this type of software in all our regression analyses.

[6]The option to select out of digital surveillance is provided by a prompt from the tool itself rather than by the manager.

hour work window and were given the option to select out of the surveillance software but their performance was not used to justify the removal of the surveillance requirement (*no surveillance-unjustified*),[7] and the final third were told they had high productivity during the first two-hour window but were not given the option to select out of the surveillance software (*high-performance control*). This design allows us to test the causal effects of digital surveillance on worker performance and satisfaction, as well as how this varies by whether or not workers are given a performance-based rationale for input monitoring.

Three main findings emerge from the RCT results. First, maintaining the digital surveillance condition for low-productivity workers reduces performance relative to allowing workers to remove surveillance only when this continued surveillance is not explained relative to worker performance. Second, among high-productivity workers, explicitly removing the surveillance requirement also reduces performance relative to the control but, again, only when the removal is not explained relative to their performance. In both cases, performance drops approximately 17% relative to control workers. These patterns are consistent with the performance impacts of digital surveillance being conditional on how surveillance requirements are explained by managers. Moreover, they demonstrate that the presence or elimination of surveillance in a work environment where digital surveillance is common does not necessarily change performance, suggesting that investing in the infrastructure and human capital required to manage effective digital monitoring may not be cost effective. Rather, investing in managerial capital that workers believe is effective may be a better use of resources.

Third, we do not find consistent evidence that these performance changes are driven by differences in job satisfaction. In particular, while high-productivity workers in the no surveillance-unjustified group report being significantly less likely to accept a follow-up job than the high-productivity control, among low-productivity workers, surveillance is associated with a higher willingness to accept a follow-up job. One possible explanation for this inconsistency is that managerial weakness may be interpreted as beneficial for low-performers who can benefit from lower quality oversight, but as undesirable for high-performers who can benefit from high quality oversight. In

[7] As in the low-performance control group, the option to select out of input monitoring was provided by a prompt from the tool itself rather than by the manager.

addition, we find that high-productivity workers in the no surveillance-justified group report significantly lower satisfaction with their pay on the job than workers in both the high-performance control and the no surveillance-unjustified groups. This suggests that removing of digital surveillance as a reward for good performance is not valued relative to pay increases.

The paper contributes to two literatures. First, we contribute to the fast-developing literature on the heterogeneity in the returns to management practices. We study a management practice that is commonly used and increasing in importance, namely, digital worker surveillance. Prior work has shown that while certain management practices have large average returns (Bloom et al., 2013), the returns to management practices for firms can vary by regulatory environments and knowledge spillovers (Bloom et al., 2019), as well as by workplace culture (Blader et al., 2020). We add to this literature by showing that subtle differences in managerial justification workers receive for their surveillance environment can impact the returns to surveillance. Randomly varying both whether a worker is monitored and whether that is justified based on their performance ensures that we are able to test for these channels causally.

Second, prior work has shown that monitoring can improve worker productivity (Pierce et al., 2015; de Rochambeau, 2017; Kelley et al., 2021; Blader et al., 2020) and process compliance (Staats et al., 2017), as well as reduce undesirable workplace behaviors like shirking (Duflo et al., 2012) and theft (Pierce et al., 2015).[8] However, qualitative work shows that workers dislike and respond to surveillance with behaviors that reduce the efficacy of monitoring (Anteby and Chan, 2018; Bernstein, 2012). This is consistent with quantitative work showing a large heterogeneity in the returns to monitoring, including some studies that find overall negative effects (Bernstein, 2012). Blader et al. (2020) find that workplace culture impacts the treatment effects of monitoring on productivity, and de Rochambeau (2017) finds that certain types of workers respond negatively to monitoring.[9] We contribute to this literature by showing that the returns to monitoring can depend on managerial justification, which we vary experimentally. In particular, we find that, even within a given setting and workforce, both the removal and the continuation of digital monitoring

[8] Varying the salience of monitoring can also increase productivity (Jensen et al., 2020).

[9] Prior work has also highlighted that political economy factors like political connections can impact the returns to monitoring in the public sector (Callen et al., 2023).

can harm worker performance, demonstrating that the effects of monitoring are not straightforward and depend on how monitoring intensity is explained to workers.

2 Research Design

2.1 Research Setting

To address our research questions, we ran an RCT on UpWork, a large online labor market through which remote workers located around the world apply and work on jobs posted by employers located around the world.[10] We advertised an 8-hour data entry job on UpWork and hired all applicants who were willing to work for our maximum hourly wage of 10 USD (the minimum we were willing to pay was 3 USD)[11] until we reached our target sample size of 400 workers. The text used in our job advertisement is provided in Appendix B. Our final sample includes 434 workers because job offers for our final pool of applicants went out concurrently and we wanted to be sure that at least our target sample accepted and worked on the job.

The data entry task workers were assigned to required that they enter financial statements of public sector enterprises in India between 1992-2009 from scanned PDF files to spreadsheets. The financial statements are imperfectly scanned such that, for instance, some appear off-center and slanted and most are slightly blurry. Given this format, the performance of automatic data entry software on the task is poor and the use of manual entry is justified as a result. The accuracy of workers' output is measured relative to entries made by a professional data entry firm before the start of our RCT. The job instruction text workers receive is provided in Appendix B.

We ask workers to submit the work they have completed after an initial 2-hour work period to allow us to assess worker performance on the task.[12] During this first 2-hour window, all workers in our study undertake their work with the UpWork ''Work Diary" activated. The Work Diary program automatically tracks worker activity through screenshots and keystroke and type activity

[10] See Agrawal et al. (2015) and Burbano (2021) for more detail on the UpWork platform.

[11] We re-posted the job advertisement weekly to capture new pools of workers because job ads are frequently sorted by date of posting on the UpWork platform.

[12] Milestone work submissions are common in hourly wage contracts on UpWork because they help employers insure against paying for hours of work done incorrectly.

tracking.[13] Workers hired on hourly wage contracts are given access to the work diary to reduce disputes over the number of hours worked but employers have the option to give workers permission to manually add hours not logged while being tracked by the work diary.[14] After our sample of workers' first 2-hour output is evaluated, they are asked to continue working for 6 hours and are paid for 8 hours of work.

2.2 Experimental Design

Following the submission and evaluation of workers' output following 2-hours of work, we randomly assign workers to treatment conditions for the remaining 6-hours of work depending on their performance up to that point. Specifically, we evaluate whether worker performance measured as number of correct cell entries is high or low relative to a baseline median.[15]

High-performing workers are randomly assigned to one of three groups; *no surveillance-justified* , *no surveillance-unjustified*, and *high-performance control*. All workers in these three groups are told they had high productivity in the first two hours of work and asked to maintain this performance level for the remainder of the task, but their monitoring conditions varied. Specifically, workers in the *no surveillance-justified* group are told they earned a lack of monitoring through their performance and no longer need to perform their work while logged into the monitoring software. Workers in the *no surveillance-unjustified* group are also given the option to no longer perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. Workers in the *high-performance control* group are not given the option to perform the remainder of their work without the monitoring software. We refer to this third group as the control because they receive the same information about their performance as the other two groups

[13] A screenshot and activity summary is sent to the employer every 10 minutes that a freelancer tracks time. This tool is similar to stand-alone digital worker monitoring technologies like, for example, TopTracker and Time Doctor's employee monitoring tools, which track work time continuously through keystroke and type activity and capture screenshots across pre-specified time intervals.

[14] Workers hired on fixed-price contracts cannot be required to track their work through the Work Diary program because payment is based on output alone. Based on UpWork recommendations and a scan of jobs posted in April 2022, fixed-price contracts for data entry jobs are relatively rare.

[15] We hired 20 workers to perform 2 hours of data entry work before running the RCT to generate a baseline performance distribution. We updated the baseline distribution using the first 2 hours of work submissions from our study sample after the first month, second month, and fourth month of data collection. Our LASSO estimator includes week fixed effects to ensure that this does not impact our results.

but are not given the same removal of monitoring.[16]

Low-productivity (below median) workers are also randomly assigned to one of three groups; *surveillance-justified*, *surveillance-unjustified*, and *low-performance control*. As with the high-performing workers, all low-performing workers are given truthful information about their productivity, in that they are told they had low productivity in the first two hours of work and asked to improve their performance for the remainder of the task, but their monitoring conditions varied. Workers in both the *justified surveillance* and *unjustified surveillance* are required to continue working the remaining 6 hours of the job while logged into the Work Diary program. However, those in the *justified surveillance* are told this requirement is related to their low performance and for those in the *unjustified surveillance*, this was not tied explicitly to their performance. Workers in the *low-performance control* group within this subsample are permitted to select out of working while logged into the Work Diary. We refer to this third group as the control because they receive the same information about their performance as the other two groups but did not have to work under monitored conditions. Our treatment groups are summarized in Figure 1.

3 Data & Empirical Strategy

3.1 Data

Outcome Variables

We have three sets of main outcomes. The first is a measure of compliance, namely, whether or not workers comply with surveillance ("Surveillance"). This is a dummy variable that takes the value 1 if workers perform any of their post-treatment work while being digitally surveilled, and 0 otherwise- we verify that our interpretation of results is unchanged if we use a stricter measure

[16]Our study was carried out by three research assistants simultaneously managing hired workers. Approximately two months into the study, treatment assignment among hired workers was not uniform across treatments due to discrepancies in hiring rates across the research assistants. We updated treatment assignment rules to correct for uneven treatment assignments. Specifically, among high-performing workers, the probabilities of being assigned to no surveillance-justified, surveillance-justified, and high-performance control were increased to 39.1 percent, decreased to 18.04 percent, and increased to 42.86 percent, respectively. Among low performing workers, the probabilities of being assigned to surveillance-justified, surveillance-unjustified, and low-performance control were increased to 45 percent, increased to 35 percent, and decreased to 20 percent, respectively. We reverted back to uniform treatment assignment approximately two months later, once treatments were evenly distributed in our sample. Our LASSO estimator includes week fixed effects to ensure this variation does not impact our results.

of surveillance; whether or not workers do all of their post-treatment work while being digitally surveilled.[17] The second is worker productivity. Worker productivity on the post-treatment portion of the task is measured as the number of correct cell entries ("Number of Correct Cells Entered"),[18]. Finally, our third set of outcomes is related to job satisfaction. This is measured using post-task survey responses to questions about how much workers enjoyed the job relative to similar tasks ("Enjoyed Job"), whether they think they were fairly paid for the job ("Felt Fairly Paid"), and whether they would accept a subsequent job offer from us ("Would Accept Another Job"). The full list of post-task survey questions is provided in Appendix B.

Treatment Status

Our primary independent variables of interest are indicators for worker treatment conditions as described in Section 2.2. We estimate the effects of providing the no-surveillance option to high-performers regardless of whether this option was justified, and then separately by whether or not it was justified based on worker performance. Similarly, we estimate the effects of requiring surveillance for low-performers regardless of whether this requirement was justified, and then separately by whether or not it was justified based on worker productivity.

Baseline Worker Characteristics

We also collect data on worker characteristics from their UpWork profile pages. Specifically, we captured information on their geographic location (country), gender inferred from names and pictures, and the wage they publicly advertise as their target hourly rate. In addition, we record the week each worker was hired, which one of three research assistants hired and managed each worker, and whether a worker is suspected by their managing research assistant of using data entry software.[19] In addition, we capture workers' intrinsic and extrinsic work motivation from post-task survey questions. These ask the worker about the extent to which they work on data entry tasks because of the satisfaction they get from succeeding at difficult tasks and the extent to which they

[17] Workers could not comply with our surveillance instructions by submitting requests for payment for work done not in the Work Diary despite being assigned to a surveillance condition, or by continuing to undertake work while logged into the Work Diary despite being assigned to a no surveillance condition.

[18] Additionally, we analyze whether cell entry errors are affected by treatment and find that they are not (see Table A2, and whether total entries uncorrected for errors are affected by treatment and find similar results as we do in our main productivity analysis (see Table A3).

[19] As we report in Table A1, suspected software use is unrelated to treatment conditions. We also verify that our treatment effect estimates are unaffected by dropping workers suspected of using data entry software.

work on data entry tasks for the income they earn from them, respectively.

We report regression coefficients and sample sizes from regressions comparing baseline characteristics for high-productivity workers in Table 1 and for low-productivity workers in Table 2. The comparison groups are listed in column headings. We also present the joint F-test of variable significance for each treatment assignment possibility. Among both types of workers, we find few differences in worker characteristics by treatment and no evidence of systematic differences by treatment assignment. Moreover, we include all baseline characteristics in the LASSO specification described below, to ensure that any baseline differences correlated with treatment status are not impacting our results.

3.2 Empirical Strategy

We analyze the effects of our surveillance treatments on outcomes separately for high and low-performing workers, since they were randomized into treatments separately from each other and subject to different treatment conditions.

We begin by analyzing the effects of our treatment groups on whether or not workers opt out of digital surveillance. This analysis tests for compliance with treatment assignment among the low-performance worker group and for selection out of surveillance among the high-performance group. As we demonstrate, while, essentially all workers required to perform their tasks while being digitally surveilled comply with this condition, many workers who have the option to deactivate the digital surveillance tool do not do so. We follow this analysis with intent to treat (ITT) analyses of the managerial messaging treatments on worker performance and job satisfaction, respectively.

Our preferred approach for estimating our treatment effects uses a post-double selection (PDS) LASSO estimator (Belloni et al., 2014). The PDS LASSO selects which covariates to include in OLS treatment effect estimation based on which covariates predict both our outcome variable of interest and our treatment variables.[20] Relative to standard OLS estimates of our treatment effects, the PDS LASSO estimates are more precise but (as expected), do not change any of our findings

[20]The potential controls our PDS LASSO selects from are worker gender, hourly wage on our job, profile page advertised hourly wage, extrinsic and intrinsic motivation, country indicators, week indicators, research assistant indicators, baseline performance (i.e. correct cell count in first 2-hours of work), and whether or not a worker was suspected of using data entry software.

or interpretations.[21]

Specifically, to assess the effects of digital surveillance by whether or not it is justified based on worker performance, we estimate the following regression model separately for low-performance and high-performance workers:

$$Y_i = \alpha + \beta_{1i} + \text{SurveilTreatJustified}_i + \beta_2 \text{SurveilTreatUnjustified}_i + \mathbf{X}_i + \varepsilon_i, \tag{1}$$

where Y_i is the outcome variables of interest for worker i, $\text{SurveilTreatJustified}_i$ is an indicator for whether or not participant i received an explanation for their surveillance treatment (no surveillance among high performers, surveillance among low performers), $\text{SurveilTreatUnjustified}_i$ is an indicator for whether or not participant i did not received an explanation for their surveillance treatment (no surveillance among high performers, surveillance among low performers), and $\mathbf{X}_i$ is a vector of control variables and fixed effects the PDS LASSO routine selects from the list we provide it.

Motivated by existing evidence of significant heterogeneity in how workers respond to managerial monitoring (de Rochambeau, 2017; Friebel et al., 2024), we pre-registered several heterogeneity analyses and report the results of these analyses in Appendix A. We do not find any systematic evidence of heterogeneous treatment effects across gender, internal task motivation, or external task motivation and, thus, do not focus on these estimates in our main analysis (see Appendix A.2).[22]

4 Results

Surveillance by Treatment

We begin by reporting our first stage results–the effects of our treatments on worker digital surveillance–in Table 3. Columns 1 and 2 report these treatment effects for high-performers and columns 3 and 4 report them for low-performers.

Column 1 demonstrates that removing the digital surveillance requirement for high-performing

[21] These are omitted for brevity but available upon request.

[22] We also pre-registered an analysis of heterogeneous treatment effects by country-level power distance measures. We do not have a sufficiently large number of countries represented in our sample to rule out that any heterogeneous treatment effects by power-distance are driven by other country-level differences and, thus, do not report these results.

workers significantly decreased the likelihood that they performed their work while being surveilled, indicating their preference to not be monitored. However, a majority of workers opted to continue working while being digitally surveilled and column 2 demonstrates that the decrease in digital surveillance usage is fully driven by high-performers who received an explanation for the option to deactivate digital surveillance. Workers in the justified no-surveillance group were about 32 percentage points less likely to work under surveillance conditions than high-performer control workers who did not have the option to work while not being surveilled. However, workers in the unjustified no surveillance group were no less likely to work under monitored conditions than high-performer control workers.

Column 3 demonstrates that a significant proportion of low-performer control workers who had the option to select out of digital surveillance did so. In particular, workers in the justified surveillance and in the unjustified surveillance (who were not given permission to be paid for work hours not tracked by the digital surveillance tool) are both approximately 10 percentage points more likely to have had their work tracked by the digital surveillance tool than workers in the low-performance control.

Combined, the estimates reported in Table 3 demonstrate near perfect compliance with the digital surveillance requirement among workers subjected to it. More interestingly, they demonstrate that workers permitted to select out of digital surveillance are much more likely to do so when this permission is justified based on their work performance. Furthermore, there is a strong first stage between treatment assignment and compliance.

Performance by Treatment

How does surveillance impact productivity, and does managerial justification affect the returns to surveillance? The ITT effects of managerial messaging about digital surveillance requirements are reported in Table 4. Columns 1 and 2 report the ITT estimates for the sample of high-performers and columns 3 and 4 report them for the sample of low-performers. Column 1 reports a negative but noisy effect of removing the digital surveillance requirement among high-performers. Column 2 demonstrates that removing the digital surveillance requirement but not justifying this change based on worker performance significantly and meaningfully decreased worker performance

relative to control workers who did not have the option of working without digitally surveillance (but otherwise were provided with the same messaging from managers). Given that this group of workers did not take advantage of the option to work while not being tracked by digital surveillance software, this performance decline is unlikely caused by a change in moral hazard. While the coefficient estimate for high-performers who received performance-based justification for the digital surveillance requirement removal is not statistically significant, it is relatively large and negative and allows us to rule out large gains from the surveillance removal even when justification is provided.

Similarly, we find a negative but noisy effect of maintaining surveillance among low-productivity workers on average. However, in symmetry with our finding among high-productivity workers, we find that maintaining surveillance without justifying why significantly reduces worker performance (column 4), by about the same percent as not justifying removing surveillance among high performing workers does (17% relative to control). We do not find a large or significant effect of maintaining surveillance among low-productivity workers when surveillance is justified based on worker performance. In sum, the productivity results point to the importance of managerial justification in driving the impacts of adopting worker monitoring technologies.

Job Satisfaction by Treatment

One possible explanation for why the removal or continuation of digital surveillance requirements reduced performance when workers did not receive a managerial justification for the change may be that this lack of justification reduced worker engagement with the task. For instance, unexplained surveillance policies may lead workers to question the capabilities of their employer and reduce their enjoyment of the task as a result. Table 5 reports the results of our analysis of the ITT effects of managerial messaging about digital surveillance on how much workers reported enjoying the job (columns 1 and 2), whether workers state they'd accept a subsequent job offer from the same employer (columns 3 and 4), and whether or not workers thought they were fairly paid for the task (columns 5 and 6). Panel A reports these estimates for the sample of high-performers and Panel B reports them for the sample of low-performers.

Consistent with workers in the no surveillance-unjustified group reducing their performance in response to treatment, we find that they are also less likely to say they would accept a follow-up job

than workers in the high-performance control- this is despite them not being differentially likely to state they were paid fairly. In contrast, while performance among workers in the no surveillance-justified group did not significantly respond to treatment, we find that they report significantly lower satisfaction with their pay on the job than workers in both the high-performance control and the no surveillance-unjustified groups. This suggests that offering more autonomy through the removal of digital surveillance as a reward for good performance is not valued relative to pay increases, and that alerting workers to a desire to reward their performance can change their perceptions about what they deserve.

Among low performers, the surveillance treatments do not change how much workers say they enjoyed the task or the extent to which they felt they were fairly paid. However, we find an increase in the likelihood workers would accept a subsequent job from us among both the unjustified surveillance and justified surveillance groups. This suggests that the drop in performance among low-productivity workers in the surveillance-unjustified group was not driven by a change in job satisfaction and, moreover, that digital surveillance does not appear to reduce worker retention.

5 Conclusion

Our study estimates the returns to worker surveillance and how clarity about managers' decisions to implement digital surveillance affects worker performance and job satisfaction. We find little evidence that digital surveillance or the lack thereof systematically affects worker performance or jobs satisfaction in a setting where this type of monitoring is common. However, when managers do not explain the maintenance or removal of digital surveillance, worker performance falls, indicating the importance of managerial justification in determing the returns to such technologies.

Broadly, our results seem most consistent with the interpretation that worker perceptions about managerial capabilities drive performance more than monitoring itself does. Among low performers, maintaining monitoring consistency without justifying the use of digital surveillance reduces performance but maintaining this surveillance with a performance-based explanation does not. Our results among high performers demonstrate a symmetric pattern with the removal of

digital surveillance–changing monitoring conditions without justification reduces performance, and justifying the change reduces worker perceptions about the value of the job. This implication is consistent with other empirical examinations of input monitoring. For instance, Jensen et al. (2020) find significantly positive effects of input monitoring on worker performance when it increases worker perceptions of managerial productivity. Similarly, Pierce et al. (2015) find that digital employee theft monitoring both reduces theft and improves productivity due in part to changes in worker perceptions about managerial oversight. Combined with existing empirical studies, our study further supports the conclusion that whether or not monitoring improves firm performance depends on what it signals to workers about managerial capabilities.

Interestingly, we do not find consistent evidence that performance declines due to unjustified surveillance conditions are due to changes in worker satisfaction with the job. While the unjustified removal of digital surveillance among high-performers is associated with a reduction in retention likelihood, the unjustified maintenance of this surveillance among low-performers is associated with an increased retention likelihood. One possible explanation for this inconsistency is that managerial weakness may be interpreted as beneficial for low-performers who can benefit from lower quality oversight, but as undesirable for high-performers who can benefit from high quality oversight.

Our experimental design allows us to isolate the causal effects of changes in digital surveillance and how this varies by whether the use or removal of digital surveillance is justified based on worker performance, in a setting where digital surveillance is common. While our findings that managerial decision-making clarity impacts worker performance and that digital surveillance does not invariably change worker performance are important general takeaways from our study, several related questions would be complementary to study in future work. For instance, subsequent work may consider how introducing digital surveillance in settings where this type of monitoring is novel or in settings where high levels of shirking have been observed manually can expand our understanding of the conditions under which digital surveillance can enhance or impede performance.

References

Agrawal, Ajay, John Horton, Nicola Lacetera, and Elizabeth Lyons, "Digitization and the contract labor market," *Economic analysis of the digital economy*, 2015, *219*, 89–110.

_ , Nicola Lacetera, and Elizabeth Lyons, "Does standardized information in online markets disproportionately benefit job applicants from less developed countries?," *Journal of international Economics*, 2016, *103*, 1–12.

Anteby, Michel and Curtis K Chan, "A self-fulfilling cycle of coercive surveillance: Workers' invisibility practices and managerial justification," *Organization Science*, 2018, *29* (2), 247–263.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 2014, *81* (2), 608–650.

Bernstein, Ethan S, "The transparency paradox: A role for privacy in organizational learning and operational control," *Administrative Science Quarterly*, 2012, *57* (2), 181–216.

Blader, Steven, Claudine Gartenberg, and Andrea Prat, "The contingent effect of management practices," *The Review of Economic Studies*, 2020, *87* (2), 721–749.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does management matter? Evidence from India," *The Quarterly journal of economics*, 2013, *128* (1), 1–51.

_ , Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen, "What drives differences in management practices?," *American Economic Review*, 2019, *109* (5), 1648–1683.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Burbano, Vanessa C, "The demotivating effects of communicating a social-political stance: Field experimental evidence from an online labor market platform," *Management Science*, 2021, *67* (2), 1004–1025.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee, "The political economy of public sector absence," *Journal of Public Economics*, 2023, *218*, 104787.

de Rochambeau, Golvine, "Monitoring and intrinsic motivation: Evidence from Liberia's trucking firms," *Unpublished Manuscript*, 2017.

Duflo, Esther, Rema Hanna, and Stephen P Ryan, "Incentives work: Getting teachers to come to school," *American economic review*, 2012, *102* (4), 1241–1278.

Frey, Bruno S, "Shirking or work morale?: The impact of regulating," *European Economic Review*, 1993, *37* (8), 1523–1532.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, Tobias Kretschmer, and Nick Zubanov, "Is this really kneaded? Identifying and eliminating potentially harmful forms of workplace control," Technical Report, ECONtribute Discussion Paper 2024.

Gambardella, Alfonso, Pooyan Khashabi, and Claudio Panico, "Managing autonomy in industrial research and development: A project-level investigation," *Organization Science*, 2020, *31* (1), 165–181.

jen Chang, Juin and Ching chong Lai, "Carrots or sticks? A social custom viewpoint on worker effort," *European Journal of political economy*, 1999, *15* (2), 297–310.

Jensen, Nathan, Elizabeth Lyons, Eddy Chebelyon, Ronan Le Bras, and Carla Gomes, "Conspicuous monitoring and remote work," *Journal of Economic Behavior & Organization*, 2020, *176*, 489–511.

Kelley, Erin M, Gregory Lane, and David Schönholzer, "Monitoring in target contracts: Theory and experiment in Kenyan public transit," 2021.

Ko, Yea Hee and In Gyun Baek, "The Effect of Computer Monitoring on Employees' Productivity in Telecommuting Arrangements," *Management Science*, 2024.

Pierce, Lamar, Daniel C Snow, and Andrew McAfee, "Cleaning house: The impact of information technology monitoring on employee theft and productivity," *Management Science*, 2015, *61* (10), 2299–2319.

Staats, Bradley R, Hengchen Dai, David Hofmann, and Katherine L Milkman, "Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare," *Management Science*, 2017, *63* (5), 1563–1585.

Tables and Figures

Figure 1: Experimental Design

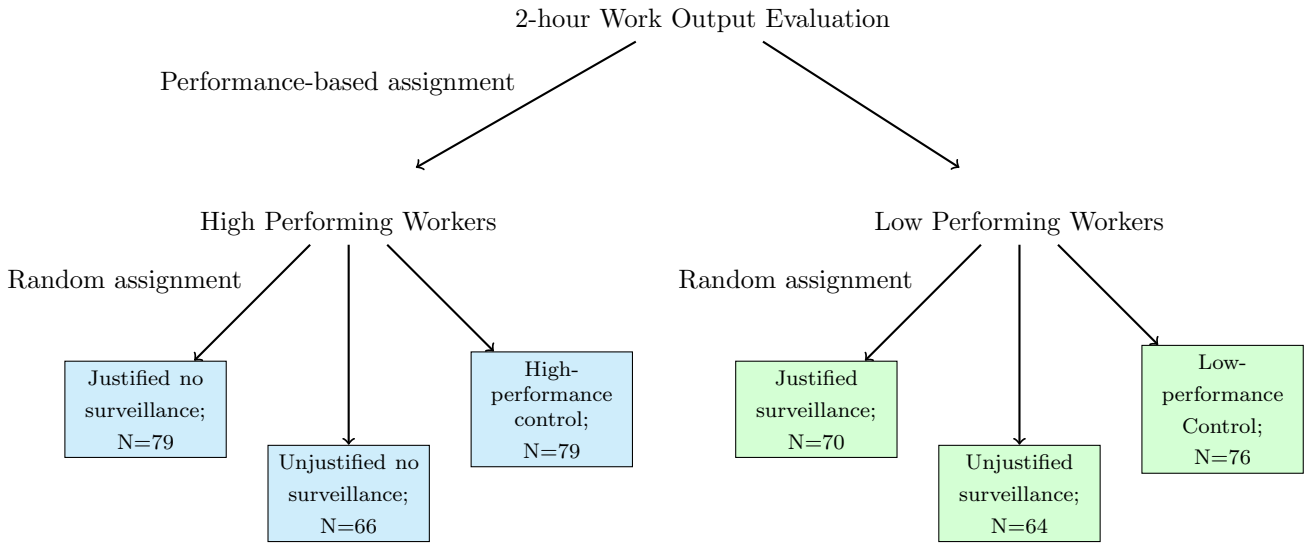


Table 1: Worker Characteristics by Treatment, High Productivity Types

	No Surveillance vs. Control		Justified No Surveillance vs. Control		Unjustified No Surveillance vs. Control		Justified No Surveillance vs. Unjustified No Surveillance	
	Coefficient	N	Coefficient	N	Coefficient	N	Coefficient	N
# of Correct Cells	-401.7	221	-443.2	155	-351.9	142	-91.30	145
	(293.7)		(303.5)		(315.6)		(189.6)	
Correct % Rate	-0.0003	221	0.00392	154	-0.00544	142	0.00937*	145
	(0.004)		(0.00447)		(0.00541)		(0.00511)	
Bid Wage	0.341	220	0.392	154	0.280	141	0.111	145
	(0.309)		(0.360)		(0.366)		(0.378)	
Male	0.104	221	0.123	155	0.0809	142	0.0418	145
	(0.0707)		(0.0802)		(0.0844)		(0.0837)	
Advertised Wage	-1.090	221	-0.596	155	-1.681**	142	1.085	145
	(0.726)		(0.0801)		(0.755)		(0.853)	
High Ext Motivation	0.0999	221	0.138*	155	0.0538	142	0.0846	145
	(0.0690)		(0.0791)		(0.0823)		(0.0832)	
High Int Motivation	0.0211	221	0.0160	155	0.0271	142	-0.0111	145
	(0.0701)		(0.0796)		(0.0832)		(0.0822)	
P-Value for Joint Test		0.12		0.12		0.15		0.33

Notes: This table presents regression coefficients and sample sizes from regressions comparing baseline characteristics of high-productivity workers. The comparison groups are listed in column headings. # of Correct Cells is the primary productivity measure, and is the number of correct entries made by the worker in the data entry task, bench-marked using the same data entered by a professional services firm. The Correct % Rate is a related and secondary productivity measure, which measures the error rate in the data entered by the worker. Bid Wage is the hourly wage in USD that workers bid for the project. Male is a dummy variable that takes the value 1 if the worker presented as male as measured by their name and photograph, and 0 otherwise. Advertised wage is the worker's wage listed on their profile. High Ext Motivation is a dummy variable that takes the value 1 if the worker reported above median levels of agreement with the survey question "I do data entry work for the income it provides me.", of seven possible levels of agreement. High Int Motivation is a dummy variable that takes the value 1 if the worker reported above median levels of agreement with the survey question "I do data entry work for the satisfaction I experience when I am successful at doing difficult tasks.", of seven possible levels of agreement. The full survey is included in the Data Appendix. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after the baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions.

Table 2: Worker Characteristics by Treatment, Low Productivity Types

	Surveillance vs. Control		Justified Surveillance vs. Control		Unjustified Surveillance vs. Control		Justified Surveillance vs. Unjustified Surveillance	
	Coefficient	N	Coefficient	N	Coefficient	N	Coefficient	N
# of Correct Cells	-97.87	213	-171.4*	146	-21.02	143	-150.4	137
	(94.08)		(93.68)		(118.0)		(99.03)	
Correct % Rate	0.0134	213	0.003	146	0.0243	143	-0.0213	137
	(0.0166)		(0.0210)		(0.0161)		(0.0173)	
Bid Wage	0.523*	211	0.675*	144	0.366	142	-0.0213	136
	(0.294)		(0.358)		(0.329)		(0.356)	
Male	-0.0548	213	-0.0658	146	-0.0434	143	-0.0224	137
	(0.0714)		(0.0831)		(0.0840)		(0.0860)	
Advertised Wage	-0.961	213	-0.795	146	-1.681**	143	0.337	137
	(0.905)		(0.984)		(0.953)		(0.683)	
High Ext Motivation	-0.0490	213	-0.0955	146	0.0538	143	-0.0951	137
	(0.0716)		(0.0830)		(0.0839)		(0.0857)	
High Int Motivation	-0.109	213	-0.105	146	-0.113	143	0.00746	137
	(0.0708)		(0.0825)		(0.0835)		(0.0861)	
P-Value for Joint Test		0.16		0.20		0.27		0.36

Notes: This table presents regression coefficients and sample sizes from regressions comparing baseline characteristics for low-productivity workers. The comparison groups are listed in column headings. # of Correct Cells is the primary productivity measure, and is the number of correct entries made by the worker in the data entry task, bench-marked using the same data entered by a professional services firm. The Correct % Rate is a related and secondary productivity measure, which measures the error rate in the data entered by the worker. Bid Wage is the hourly wage in USD that workers bid for the project. Male is a dummy variable that takes the value 1 if the worker presented as male as measured by their name and photograph, and 0 otherwise. High Wage is a dummy variable that takes the value 1 if the worker's bid wage is greater than the median value. Advertised wage is the worker's wage listed on their profile. High Ext Motivation is a dummy variable that takes the value 1 if the worker reported above median levels of agreement with the survey question "I do data entry work for the income it provides me.", of seven possible levels of agreement. High Int Motivation is a dummy variable that takes the value 1 if the worker reported above median levels of agreement with the survey question "I do data entry work for the satisfaction I experience when I am successful at doing difficult tasks.", of seven possible levels of agreement. The full survey is included in the Data Appendix. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Justified Surveillance group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Unjustified Surveillance group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Table 3: Compliance

	Worked Under Surveillance			
	(1)	(2)	(3)	(4)
No Surveillance	-0.196***			
	(0.0429)			
No Surveillance-Justified		-0.321***		
		(0.0543)		
No Surveillance-Unjustified		-0.00275		
		(0.0237)		
Surveillance			0.119***	
			(0.0409)	
Surveillance-Justified				0.118***
				(0.0408)
Surveillance-Unjustified				0.119***
				(0.0404)
Baseline Productivity	High	High	Low	Low
Control Mean	1	1	.84	.84
P-Val: Justified= Unjustified		0.000***		0.963
N	219	219	208	208

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. Worked Under Surveillance is a dummy variable that takes the value 1 if the worker chose to continue working under monitoring conditions after baseline measurement, and 0 otherwise. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The No Surveillance-Justified group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer need to perform their work while logged into the monitoring software. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Table 4: Productivity

	Productivity (Number of Correct Cells Entered)			
	(1)	(2)	(3)	(4)
No Surveillance	-727.5			
	(725.4)			
No Surveillance-Justified		-871.1		
		(786.8)		
No Surveillance-Unjustified		-1729.7**		
		(726.9)		
Surveillance			-836.4	
			(701.4)	
Surveillance-Justified				-257.9
				(767.8)
Surveillance-Unjustified				-1461.9*
				(807.5)
Baseline Productivity	High	High	Low	Low
Control Mean	10136.18	10136.18	8357.76	8357.76
P-Val: Justified= Unjustified		0.258		0.118
N	212	212	197	197

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Table 5: Job Satisfaction

Panel A: High Baseline Productivity						
	(1)	(2)	(3)	(4)	(5)	(6)
	Enjoyed Job		Would Accept Another Job		Felt Fairly Paid	
No Surveillance	-0.317		-0.150		-0.286**	
	(0.303)		(0.104)		(0.129)	
No Surveillance-Justified		-0.184		-0.113		-0.326**
		(0.314)		(0.108)		(0.134)
No Surveillance-Unjustified		-0.293		-0.217*		-0.00169
		(0.333)		(0.114)		(0.141)
Control Mean	5.351	5.351	6.929	6.929	6.579	6.579
P-Val: Justified= Unjustified		0.735		0.344		0.018**
N	175	175	173	173	175	175
Panel B: Low Baseline Productivity						
	(1)	(2)	(3)	(4)	(5)	(6)
	Enjoyed Job		Would Accept Another Job		Felt Fairly Paid	
Surveillance	0.0430		0.139		0.181	
	(0.291)		(0.0898)		(0.208)	
Surveillance-Justified		0.0153		0.209*		0.0818
		(0.373)		(0.115)		(0.267)
Surveillance-Unjustified		0.133		0.172*		0.205
		(0.340)		(0.103)		(0.242)
Control Mean	5.262	5.262	6.75	6.75	0	0
P-Val: Justified= Unjustified		0.755		0.750		0.649
N	156	156	152	152	155	155

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. Enjoyed Job is the level of agreement (out of seven possible agreement levels) with the statement "On a scale of 1-7, how much did you enjoy working on this text as compared to similar tasks you have worked on." Would accept another job is the level of agreement (out of seven possible agreement levels) with the statement "On a scale of 1-7, how likely would you be to accept another data entry task from us?". Felt Fairly Paid is the level of agreement (out of seven possible agreement levels) with the statement "I feel like I was paid fairly for this job.". The full survey is included in the Data Appendix. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Appendix A Appendix Tables

Appendix A.1 Additional Outcomes

Table A1: Suspected Non-Manual Data Entry by Treatment

	Suspected Non-Manual Entry			
	(1)	(2)	(3)	(4)
No Surveillance	-0.0587			
	(0.132)			
No Surveillance-Justified		-0.0847		
		(0.128)		
No Surveillance-Unjustified		-0.105		
		(0.133)		
Surveillance			-0.0907	
			(0.104)	
Surveillance-Justified				-0.0712
				(0.0975)
Surveillance-Unjustified				-0.0415
				(0.119)
Baseline Productivity	High	High	Low	Low
Control Mean	.49	.49	.34	.34
P-Val: Justified= Unjustified		0.863		0.749
N	219	219	208	208

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. Suspected Non-Manual Data Entry is a dummy variable takes the value 1 if the worker may have used data entry software, and 0 otherwise. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Table A2: Error Rate by Treatment

	Error Rate			
	(1)	(2)	(3)	(4)
No Surveillance	0.00604			
	(0.0180)			
No Surveillance-Justified		-0.00600		
		(0.0210)		
No Surveillance-Unjustified		0.00211		
		(0.0190)		
Surveillance			-0.0174	
			(0.0183)	
Surveillance-Justified				-0.0221
				(0.0210)
Surveillance-Unjustified				-0.0172
				(0.0206)
Baseline Productivity	High	High	Low	Low
Control Mean	.92	.92	.87	.87
P-Val: Justified= Unjustified		0.657		0.819
N	212	212	197	197

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. Error Rate is defined as the number of correct cells entered, divided by the number of total cells entered by the worker. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Table A3: Total Number of Entries by Treatment

	Total Output			
	(1)	(2)	(3)	(4)
No Surveillance	-690.7			
	(769.2)			
No Surveillance-Justified		-682.5		
		(833.8)		
No Surveillance-Unjustified		-1846.5**		
		(756.3)		
Surveillance			-1231.1	
			(827.1)	
Surveillance-Justified				-588.6
				(923.3)
Surveillance-Unjustified				-1755.4*
				(975.1)
Baseline Productivity	High	High	Low	Low
Control Mean	11130.78	11130.78	9813.24	9813.24
P-Val: Justified= Unjustified		0.148		0.205
N	212	212	197	197

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. Total Number of Entries is the number of total cells entered by the worker, regardless of whether they are correct or not. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 1 and 2 is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Surveillance-Justified group is the subset of the low-productivity workers who were randomly told that they had earned continued monitoring through their (low) performance and needed to perform their work while logged into the monitoring software. The Surveillance-Unjustified group is the subset of the low-productivity workers who were randomly told that they needed to continue performing their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Appendix A.2 Pre-Registered Heterogeneity Analyses

Table A4: Heterogeneous Treatment Effects by Gender, Internal Motivation, and External Motivation: High Productivity Workers

	Productivity (Number of Correct Cells Entered)					
	(1)	(2)	(3)	(4)	(5)	(6)
No Surveillance X Het. Var.	1450.9					
	(1274.2)					
Het. Var.	281.6	217.4				
	(1056.8)	(1059.7)				
No Surveillance	-1506.8		348.3		-1020.8	
	(989.2)		(1081.7)		(951.2)	
No Surveillance-Unjustified X Het. Var.		1884.7				
		(1484.6)				
No Surveillance-Justified X Het. Var.		1129.6				
		(1639.5)				
No Surveillance-Unjustified		-2701.5***		-1113.3		-2199.9**
		(1036.5)		(1229.5)		(1004.0)
No Surveillance-Justified		-1528.0		332.8		-1586.3
		(1151.6)		(1385.5)		(1042.2)
No Surveillance X Het. Var.			-1316.1			
			(1320.0)			
Het. Var.			2916.4**	2977.1**		
			(1194.4)	(1239.1)		
No Surveillance-Unjustified X Het. Var.				-1084.9		
				(1548.1)		
No Surveillance-Justified X Het. Var.				-2094.5		
				(1677.9)		
No Surveillance X Het. Var.					1027.0	
					(1418.7)	
Het. Var.					-1547.5	-1465.8
					(1237.7)	(1155.3)
No Surveillance-Unjustified X Het. Var.						774.6
						(1572.9)
No Surveillance-Justified X Het. Var.						1208.7
						(1634.9)
Het. Var.	Gender	Gender	Internal Motivation	Internal Motivation	external Motivation	external Motivation
Baseline Productivity	High	High	High	High	High	High
Control Mean	10136.18	10136.18	10136.18	10136.18	10136.18	10136.18
P-Val: Justified X Het. Var.= Unjustified X Het. Var.		0.624		0.521		0.787
N	212	212	212	212	212	212

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. The No Surveillance group is the group of high-productivity (above median baseline productivity workers) that randomly had the requirement to work under monitored conditions removed after baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group is the subset of the high-productivity workers who were only given their productivity information, and continued to work under monitored conditions.

Table A5: Heterogeneous Treatment Effects by Gender, Internal Motivation, and External Motivation: Low Productivity Workers

	Productivity (Number of Correct Cells Entered)					
	(1)	(2)	(3)	(4)	(5)	(6)
Surveillance X Het. Var.	-486.1					
	(1337.0)					
Het. Var.	-285.5	-328.6				
	(1108.8)	(1169.4)				
Surveillance	-401.7		-358.4		-1535.6	
	(1024.7)		(1001.5)		(1015.5)	
Surveillance-Unjustified X Het. Var.		602.2				
		(1518.0)				
Surveillance-Justified X Het. Var.		-765.2				
		(1633.2)				
Surveillance-Unjustified		-1724.2		-1365.0		-3485.0***
		(1135.1)		(1185.5)		(1222.3)
Surveillance-Justified		67.29		909.0		-881.4
		(1171.9)		(1273.9)		(1188.5)
Surveillance X Het. Var.			-508.1			
			(1359.2)			
Het. Var.			1613.6	1949.0		
			(1131.4)	(1238.7)		
Surveillance-Unjustified X Het. Var.				378.9		
				(1682.1)		
Surveillance-Justified X Het. Var.				-2010.6		
				(1719.5)		
Surveillance X Het. Var.					1636.5	
					(1419.3)	
Het. Var.					-629.3	-1092.6
					(1126.8)	(1196.1)
Surveillance-Unjustified X Het. Var.						3199.8**
						(1617.8)
Surveillance-Justified X Het. Var.						1274.6
						(1613.7)
Het. Var.	Gender	Gender	Internal Motivation	Internal Motivation	external Motivation	external Motivation
Baseline Productivity	Low	Low	Low	Low	Low	Low
Control Mean	8357.76	8357.76	8357.76	8357.76	8357.76	8357.76
P-Val: Justified X Het. Var.= Unjustified X Het. Var.		0.394		0.156		0.227
N	197	197	197	197	197	197

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. The Surveillance group is the group of low-productivity (below median baseline productivity workers) that randomly had to continue working under monitored conditions after the baseline measurement. The Justified No Surveillance group is the subset of the high-productivity workers who were randomly told that they had earned lack of monitoring through their performance and no longer needed to perform their work while logged into the monitoring software. The Unjustified No Surveillance group is the subset of the high-productivity workers who were randomly told that they no longer needed to perform their work while logged into the monitoring software but this option is not explicitly tied to their performance. The Control group in Columns 3 and 4 is the subset of the low-productivity workers who were only given their productivity information, and had the requirement to work under monitored conditions removed after baseline measurement.

Appendix B Data Appendix

Appendix B.1 Job Advertisement

Below is the text and information in the job advertisement posted on UpWork to recruit workers for our study.

Job Title: Data Entry from PDFs into Excel

Job Category: Data entry

Job Skills: Data entry, Microsoft Excel

Job Description: We would like assistance entering financial statement data from pdf files into excel. The task will take no more than 8 hours.

Budget: $3-10/hour

Appendix B.2 Job Instructions

Below is the text we provided hired workers on completing the task we hired them for.

Thank you for applying to this job.

We would like you to enter financial data provided in the attached pdf document into an excel spreadsheet. We would like you to enter each financial data point in a separate row. In particular, each spreadsheet row should have the company name, year, document title (e.g. balance sheet, profit and loss, management ratio, etc.), pdf sub-headers (e.g. liabilities, loans from foreign parties, etc), and the corresponding amount in Lakhs. We have attached an example spreadsheet here with the first two lines entered from the 1994-1995 page for your reference.

Please keep the column titles as they are in the attached sample spreadsheet, and make sure you are entering the data in the order it shows up in the pdfs.

Please do not share the pdfs with anyone.

We have noticed that pdf scraping softwares make many errors in inputting the data from the PDF into Excel. Please do not use a software to input the information into Excel. We are hiring you to

do the data entry manually.

We would like you to submit the data you have entered after 2 hours of work so we can assess your progress before you continue with the remaining work. We have restricted your weekly hours to 2 but we will increase this after the milestone. Please submit your work with all years of data on the same sheet in one Excel file and be sure that every column is filled out in each row. Please complete your work while logged into the UpWork work diary to make sure that your hours worked are recorded for payment. If you're unsure how to log into the Work Diary, please review UpWork's instructions here (https://support.upwork.com/hc/en-us/articles/211068518-Use-Your-Work-Diary).

Let us know if you have any questions. Please complete your work independently. Can you confirm that you won't use an automated software to do the task? Can you confirm that you will complete your work using the time tracker? We will not pay for hours not tracked in the Work Diary.

Thanks!

Appendix B.3 Treatment Assignment Text

Below is the text we provided hired workers after their 2-hour work output had been evaluated by the treatment group they were assigned to. All groups receive the standard job information in addition to the treatment-specific information.

Justified No Monitoring Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and observed high data entry productivity (number and percent of correct entries). We are hoping that you will maintain this high quality in the remaining work. As a result of your great performance, you do not need to perform the remainder of the data entry task using the UpWork App.

- *Work diary requirement is turned off*

Unjustified No Monitoring Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and observed high data entry productivity (number and percent of correct entries). We are hoping that you will maintain this high quality in the remaining work.

- *Work diary requirement is turned off*

High-Performance Control Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and observed high data entry productivity (number and percent of correct entries). We are hoping that you will maintain this high quality in the remaining work.

- *Work diary requirement is left on*

Justified Monitoring Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and noticed the productivity (number and percent of correct entries) in your data entry is low. We are hoping that you can increase your productivity for the remaining work. As a result of your lower productivity, we would like you to perform the remainder of the data entry task using the UpWork App.

- *Work diary requirement is left on*

Unjustified Monitoring Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and noticed the productivity (number and percent of correct entries) in your data entry is low. We are hoping that you can increase your productivity for the remaining work.

- *Work diary requirement is left on*

Low-Performance Control Group Information:

Thank you for submitting your first milestone work. To assess your performance on this assignment, we checked a random subset of your entered data and noticed the productivity (number and percent of correct entries) in your data entry is low. We are hoping that you can increase your productivity for the remaining work.

- *Work diary requirement is turned off*

Standard Job Information

Please enter as much of the remaining information in the pdf as possible within the remaining 6 hours of the job. We also send some additional data along in case you can finish what we had sent you within the time limit, but we realize completing it all within this time may not be possible. Please do not enter the additional data until you have finished the files sent in the first batch of instructions. Please start a new spreadsheet for the rest of your data entry.

Appendix B.4 Post-Task Survey

After submitting their final work output, workers received the following invitation to complete the post-task survey:

Thank you for submitting your work! We are hoping to understand more about how to effectively manage on UpWork. Would you be willing to answer several questions about your experience on this job for a $1 bonus? The link is here. Your willingness to answer the questions and your answers themselves will have no impact on how we review your work.

The post-task survey questions in the order that they were asked are below.

1. Please tell us your name as it appears on your UpWork profile
2. On a scale of 1-7, how much did you enjoy working on this text as compared to similar tasks you have worked on?
   (a) This was the least enjoyable task that I have ever worked on
   (b) This task was much less enjoyable than other data entry tasks I have done
   (c) This task was somewhat less enjoyable than other data entry tasks I have done
   (d) This task was as enjoyable as other data entry tasks I have done
   (e) This task was somewhat more enjoyable than other data entry tasks I have done
   (f) This task was much more enjoyable as other data entry tasks I have done
   (g) This was the most enjoyable task that I have ever worked on

   For the next two questions, please tell us how much you agree with the following statements on a scale of 1-7 where 1 is completely disagree and 7 is completely agree.
3. I do data entry work for the satisfaction I experience when I am successful at doing difficult tasks
   (a) Completely disagree
   (b) Mostly disagree
   (c) Somewhat disagree
   (d) Neither agree nor disagree

(e) Somewhat agree

(f) Mostly agree

(g) Completely agree

4. I do data entry work for the income it provides me

   (a) Completely disagree

   (b) Mostly disagree

   (c) Somewhat disagree

   (d) Neither agree nor disagree

   (e) Somewhat agree

   (f) Mostly agree

   (g) Completely agree

   For the next two questions, please tell us how accurate the statements are on a scale of 1-7 where 1 is not at all true and 7 is completely true.

5. I feel like I was treated with respect on this job

   (a) Not at all true

   (b) Mostly untrue

   (c) Somewhat untrue

   (d) Neither true nor untrue

   (e) Somewhat true

   (f) Mostly true

   (g) Completely true

6. I feel like I was paid fairly for this job

   (a) Not at all true

   (b) Mostly untrue

   (c) Somewhat untrue

   (d) Neither true nor untrue

   (e) Somewhat true

   (f) Mostly true

   (g) Completely true

7. How many hours did you spend on the data entry task?

   (a) Less than 3 hours

   (b) Between 3 and 4 hours

   (c) Between 4 and 5 hours

(d) Between 5 and 6 hours
(e) Between 6 and 7 hours
(f) Between 7 and 8 hours
(g) Between 8 and 9 hours
(h) Between 9 and 10 hours
(i) More than 10 hours

8. On a scale of 1-7, how likely would you be to accept another data entry task from us?

(a) I would definitely not accept another job from you
(b) I would almost surely not accept another job from you
(c) I am unlikely to accept another job from you
(d) I may accept another job from you
(e) I would likely accept another job from you
(f) I would almost surely accept another job from you
(g) I would definitely accept another job from you

9. Please tell us what we could have done better to support you in your work for us