Hemming and Hawing over Hawthorne: Work Complexity and the Divergent Effects of Monitoring on Productivity

Aruna Ranganathan  
Stanford University  
Alan Benson*  
University of Minnesota

Does monitoring workers improve or impair their productivity? Existing studies offer conflicting predictions. Using personnel and operational data from an Indian garment manufacturing plant, we examine how an RFID monitoring intervention on three of the plant’s twelve production lines affected productivity. We find that the effect of monitoring varied by the complexity of the work performed. Using variation in work complexity both across and within production lines, we find that monitoring significantly increased productivity for simple work but significantly decreased productivity for complex work. We contribute to research on monitoring and productivity by demonstrating how key job characteristics that make work meaningful, such as complexity, can moderate the effect of monitoring on productivity by affecting workers’ intrinsic motivation. Results also suggest that not only does the Hawthorne effect exist, but its direction can be positive, negative, or neutral depending on work complexity.

Key words: Monitoring, productivity, information technology

The extent to which an employer monitors its workers is at the heart of the economic, social, and psychological contract that defines the employment relationship (e.g. Alchian

* Aruna Ranganathan is an assistant professor at Stanford University. Alan Benson is an assistant professor at the University of Minnesota. For helpful comments, we thank Ethan Bernstein, Hengchen Dai, Ching Ren, Jesper Sorenson, Ezra Zuckerman, and participants at the Wharton People & Organizations Conference, the Wharton People Analytics Conference, the Organization Science Winter Conference, Olin Business School, Stanford GSB, and Carlson School of Management. The usual disclaimer applies. Correspondence: arunar@stanford.edu.
Monitoring appears to be so consequential that scholars have known since at least the Hawthorne experiments that the act of monitoring itself can affect workers’ productivity, providing both a theoretical and methodological impetus for research on the effect of monitoring on productivity (Mayo 1933). Not surprisingly, then, advances in information technology that have facilitated employers’ ability to monitor their workers have had a profound impact on every aspect of the employment relationship. Recent research has showcased how technology-enabled monitoring has shaped work organization and productivity among truck drivers (e.g. Hubbard 2000, Gray and Silbey 2014, Blader et al. 2016), restaurant servers (Pierce et al. 2015), teachers (Duflo et al. 2012), public administrators (Olken 2007), hospital workers (Staats et al. 2017), physicians (Simeone and Holland 2006), and many other jobs that have traditionally enjoyed considerable autonomy.

In general, monitoring has been heralded as an effective management practice: Peter Drucker famously quipped that “what’s measured, improves,” and many academic studies have found that monitoring enhances performance either in itself or as part of a system of complementary management practices (see, for example, Aral et al. 2012, Brynjolfsson and Hitt 2000, Basker 2012, Bloom and Van Reenen 2006, Bloom et al. 2013). However, other studies caution that monitoring can be harmful to workers’ productivity. For example, monitoring workers might interfere with their privacy and deter them from productive forms of experimentation and rule-breaking (Bernstein 2012). The act of monitoring workers may also demotivate workers by impinging on their sense of autonomy (Deci and Ryan 1975), or lead them to focus on the performance measure at the expense of other productive activities (Kerr 1975, H"{o}lstrom and Milgrom 1991). Given that well-executed studies have found conflicting results, we argue that instead of debating whether monitoring generally has a net positive or negative effect on worker productivity, a more appropriate question might be: under what conditions does monitoring increase or decrease worker productivity, and why?

To shed light on this question, we investigate how monitoring affected productivity at a large garment manufacturing plant producing pants and jackets in India. The
monitoring intervention was implemented in three of twelve production lines in late 2012. The treated lines in this case were chosen for a reason unrelated to their productivity: they were closest to the technical support offices, which aided installation and maintenance of the monitoring technology. The intervention involved tagging work-in-progress garments with radio frequency identification (RFID) tags; workers were instructed to scan these tags on newly installed RFID scanners at their workstation prior to working on the garment, allowing the plant and the workers themselves to monitor individual productivity in real-time. In this context, we collected three sources of data: (1) six years of daily line-level productivity before and after the intervention for all jacket and pant lines, (2) daily operation-level productivity for a sample of days before and after the intervention for pant lines, and (3) qualitative field data.

Three key features of this setting help us shed light on our research question. First, we have detailed line-level and operation-level data before and after the intervention, for both treated and nontreated lines. Such pre- and postintervention field data on productivity are rare, since the existence of such data is typically a by-product of monitoring. It is also rare to have complete and simultaneous data on nontreated groups, which allow us to compare productivity for both treated and nontreated groups at the same point in time. Second, the monitoring intervention occurred largely in absence of any simultaneous practices that would typically accompany a monitoring technology: for treated lines, monitoring was neither tied to rewards (such as piecerates) nor punishments (such as terminations). Lastly, the field data color our understanding of the intervention and along with the institutional features of the plant, allow us to speak to theoretical mechanisms.

To preview our findings, when we pool estimates for the effect of monitoring on productivity across all treated lines, we find that monitoring had no substantially or statistically significant effect on productivity. However, motivated by field observation, we further examine whether the effect of monitoring depends on the complexity of the work being performed (here, the mean time expected to complete operations). We do so by exploiting variation in work complexity both across and within lines. First, we exploit variation in complexity between pant and jacket lines, where the average operation in
pant lines is rated as being less complex (taking less time to perform) than the average operation in jacket lines. We find that monitoring significantly increased productivity in pant lines and significantly decreased productivity in jacket lines, and that this difference is statistically significant. Second, in case we were worried that pant and jacket lines differ on dimensions other than complexity, we additionally exploit variation in work complexity across operations within pant lines. Here too, we see productivity rise for relatively simple work and fall for complex work after the monitoring intervention. To explain these findings, we suggest that, in our setting, monitoring affected productivity by increasing workers’ intrinsic motivation through gamification in the case of simple work and crowding out workers’ intrinsic motivation in the case of complex work.

This paper makes four important contributions to the study of monitoring and worker productivity. First, we shift the debate from whether monitoring affects worker productivity to the conditions under which monitoring increases or decreases productivity. Second, we produce relatively clean evidence from the field that work complexity moderates the direct effect of monitoring on worker productivity. Third, though we leave the task of affirming micromechanisms to future research, the institutional details of our setting and our field work suggest that the effect of monitoring on productivity need not rely on extrinsic motives like retribution by management or pay for performance. And finally, we demonstrate that the Hawthorne effect exists and that its direction can be positive, negative, or neutral depending on work complexity. As such, our results invite future research into how monitoring affects the employment relationship, how it depends on the institutional context, and how it interacts with a variety of other management practices.

1. Monitoring and Worker Productivity: The Case For and Against Monitoring

The relationship between monitoring and productivity has long been a central question in the management literature. Taylor’s (1914) scientific management movement adopted the premise that management must engage in monitoring to ensure that workers perform
job duties in the “one best way,” while Mayo (1930) and the human relations movement adopted the premise that workers are innately self-motivated but enjoy the attention that accompanies being observed. Burawoy (1979) and subsequent labor process scholars emphasized the “production of consent” for monitoring among workers as a key objective of the firm.

Since these early studies, advancements in technology have made it significantly easier to monitor workers, prompting firms to weigh the costs and benefits of this fundamental management practice in today’s workplaces. Further, the current interest in monitoring, technology, and productivity extends beyond just the US context to industries around the world. For example, in the Indian garment industry, which is the focus of this paper, profit margins are largely governed by labor productivity, and scholars have suggested that differences in the adoption of good management practices such as monitoring explain much of the variation in productivity (Hsieh and Klenow 2009, Bloom et al. 2013).

Prior studies have been nearly universal in their praise for technology-enabled monitoring practices, at least with regard to the complementary practices that they enable. Advancements in information technologies are often cited as perhaps the principal engine of recent economic growth (Autor et al. 2003, Bresnahan 2002, Brynjolfsson and Hitt 2000, 2003, Griliches 1994). Using firm-level data from a variety of industries, Brynjolfsson and Hitt (1995) found evidence of supranormal economic returns for investments in computers. Subsequent work by Brynjolfsson and Hitt (2000, 2003) and Aral et al. (2012) found evidence from a panel of large firms that returns to computer investments grow and appear to complement organizational practices that adapt over time.

Single-firm studies have also emphasized the particular role of technology-enabled monitoring in improving productivity and work organization. Some of these studies emphasize how monitoring encourages good behaviors. For instance, Hubbard (2000) found that the advent of onboard computers improved truckers’ driving, and Staats et al. (2017) found that RFID tracking software improved hand washing compliance among
hospital workers. Monitoring can also discourage bad behaviors. Pierce et al. (2015) found that employee monitoring technology enhanced productivity and reduced theft at a restaurant chain. Outside of the US context, Duflo et al. (2012) found that installing video cameras reduced teacher absenteeism and improved students’ performance in Indian schools, while Olken (2007) found that audit warnings discouraged administrators and contractors in Indonesia from misusing funds. These empirical studies are largely consistent with a core tenet of agency theory, namely, that monitoring (or the threat of monitoring) deters workers from opportunistic behaviors such as shirking or cheating, especially when monitoring allows employers to condition rewards or punishment on workers’ output (Shapiro and Stiglitz 1984, Hölstrom 1982).

However, notable exceptions remain, and some studies have theorized and provided evidence for the negative effects of monitoring. First, monitoring may discourage workers from innovating out of fear of reprisal. In a field experiment, Bernstein (2012) found that production lines that featured an open floor plan were less productive than randomly selected treated lines that received privacy curtains. He concluded that productivity losses associated with open floor plans and transparency were largely attributable to the unwillingness of workers on nontreated lines to experiment and improve production routines. Similarly, Campbell et al. (2011) also found that tightly monitored business units learned more slowly than loosely monitored business units. Second, by directing workers to focus on performance measures, monitoring can lead workers to neglect other productive activities (Hölstrom and Milgrom 1991). Attention to quality and peer learning are classic examples of activities that can suffer when firms monitor quantity (see, for example, Freeman and Kleiner 2005), and examples of workers gaming a wider variety of performance metrics abound (see, for example, Kerr 1975, Lazear 2006). In this way, monitoring might prompt a narrow focus by employers on what a worker produces rather than on what a worker does more broadly.

Monitoring can also reduce productivity more directly through psychosocial mechanisms. For example, following Deci and Ryan’s (1975) self-determination theory, monitoring may reduce workers’ sense of autonomy and crowd out workers’ intrinsic
motivation for performing a task, a prediction that has since been widely replicated in laboratory settings.\textsuperscript{1} Prior studies have also found that the act of monitoring itself tends to be associated with the loss of autonomy and discretion over how tasks are performed (see, for example, Alchian and Demsetz 1972, Prendergast 2002, Holman et al. 2007, Batt and Colvin 2011), and that electronic monitoring reduces workers’ sense of personal control and, thereby, job performance (Stanton and Barnes-Farrell 1996).

To summarize, despite significant scholarly attention to the practice of monitoring in organizations, both the evidence and the literature are largely divided as to whether monitoring in itself improves or impairs worker productivity. In his review of the literature on transparency, Bernstein (2017, pp. 3,4) makes this divide explicit: “To date, the transparency and privacy literatures have talked past each other,” adding that, “the time is ripe for a synthesis which can offer a coherent frame to our own field’s question of how observation affects employee performance in contemporary workplaces.” Indeed, process control technologies such as those in this study are making it increasingly easy to monitor workers, and these technologies are only one of many that do so (Gilliom and Monahan 2012, Bernstein and Li 2017). Given the renewed interest in technology-enabled monitoring and the potential for significant gains to firm profitability, it is crucial to dig deeper into how monitoring affects worker productivity.

2. Setting and Hypothesis Development

To shed light on the question of whether monitoring improves or impairs productivity, we studied a monitoring intervention in a garment manufacturing plant in India. Our field observation revealed that within the same manufacturing plant, in some situations, monitoring seemed to improve worker productivity, but in other situations, monitoring seemed to impair productivity. This observation motivated us to move away from asking whether monitoring affects productivity and toward investigating the conditions under which monitoring might improve or impair productivity. In particular, we looked closely

\textsuperscript{1} For a conceptual overview with a review of laboratory studies, see Gagné and Deci (2005). In manufacturing settings, see, for example, Kruse et al. (2010) and Benson and Sajjadi (forthcoming).
at the work itself and found a job characteristic that has thus far been absent from field research on monitoring: the complexity of the work being monitored. Indeed, existing studies on monitoring and productivity have been conducted in diverse settings ranging from restaurants to high-tech manufacturing, such that there is wide variation in the complexity of the work being monitored, which might underlie the divergent results in the literature. Hence, we draw on our field observation and alternative literatures to develop our hypotheses.

2.1. Setting: Garment Plant in India

Our setting is a manufacturing plant in India that makes men’s suits. The plant features 12 production lines, 147 garment operations within the lines, and 2,212 line workers over the study period, 2009-2014. Each line consists of a set of operations: the 9 pant lines each consist of 51 pant operations, and the 3 jacket lines each consist of 96 jacket operations. Workers are assigned to a single operation on a single line and rarely move.

In 2012, the plant management introduced a monitoring technology (a shop floor control system) on a trial basis to keep track of their production and help them meet production deadlines. The CEO of the plant had seen this technology being used in manufacturing facilities on a recent visit to China and was determined to implement the same technology at his plant upon returning to India. On October 1, 2012, plant management installed a monitoring technology on two of the nine lines producing men’s pants, and on December 1, 2012, management installed the technology on one of the three lines producing men’s jackets. Figure 1 visually depicts the empirical strategy. Notably, when the technology was installed on a line, all operations within that line became monitored. Management chose these lines to be closest to the offices of the engineers responsible for maintaining the technology and in this way, the choice of treatment lines was not related to productivity or factors correlated with productivity. As shown in Table 1, the treated lines were not systematically different with regard to

\[2\] While the plan was for the plant to roll out the technology across the rest of the plant over time, because of the high cost of this technology, this roll out did not happen.
productivity or other observable worker characteristics prior to the installation of the monitoring technology.

[Figure 1]

[Table 1]

The intervention involved tagging unfinished items with RFID tags so that their progress could be tracked down the production line. RFID scanners were installed at each workstation on the treated lines, and workers were instructed to scan the RFID garment tags on their scanners prior to working on the garment. These scanners reported the number of units produced and individual efficiency as a percentage of set targets, which was updated in real time. Although performance data were now tracked at an individual level, pay remained uncommissioned and workers continued to be paid a fixed daily wage rate. In other words, the monitoring intervention in this case should be thought of as an intervention that allowed the plant to track workers’ productivity digitally and allowed workers to track themselves by seeing their “current efficiency,” in the absence of other practices that otherwise might rely on such performance data. Figure 2 shows an RFID terminal and how the terminals were installed on treated lines.

[Figure 2]

In this context, we conducted ethnographic fieldwork as well as interviews with workers and management, driven by one of the authors and a research assistant. Observation, conducted after the technology was adopted, consisted of how production was achieved in this plant, in both treated and nontreated lines, paying specific attention to how work was supervised and monitored. We analyzed this qualitative data inductively to develop our hypotheses (Glaser and Strauss 1967, Strauss and Corbin 1990).

2.2. Hypotheses: Bringing in Complexity

Our observation in this Indian manufacturing plant revealed that there was significant heterogeneity in the complexity of the work being performed by individual workers on
Ranganathan and Benson: Hemming and Hawing over Hawthorne

the shop floor, even though they were all engaged in garment production operations. For example, one worker engaged in a simple “inseam” operation said, “I do a straight stitch along the seam . . . there’s no thinking, I just need to keep the production going.” In contrast, a worker engaged in a relatively more complex “sleeve headroll attach” operation said, “I stitch pieces of fabric varying in size onto a curved piece of felt material which supports the sleeve across the shoulder—it’s an accomplishment every single time!” Operationally, we observed management taking great pains to keep track of this variation in complexity between different tasks, which they used to allocate work and measure efficiency. In addition to its operational importance, our interview data revealed that work complexity also seemed to be important for the meaning and fulfillment that workers sought from their work.

Indeed, scholars of work have long recognized that employees in a wide range of occupations, including those in low-income jobs in developing countries, desire meaning in their work (Bellah et al. 2007, Rosso et al. 2010, Ranganathan forthcoming), further arguing that when work is meaningful, workers are more likely to be intrinsically motivated and see their work as an end in itself (Wrzesniewski et al. 2017, Bunderson and Thompson 2009). Importantly, these scholars argue that key job characteristics such as work complexity can influence the meaning associated with work (Turner and Lawrence 1965, Hackman and Oldham 1980, Campbell 1988). Broadly, complexity is understood as the degree to which work in its very substance requires thought and imposes cognitive demands on the individual performing the work (Kohn and Schooler 1978). Several scholars have attempted to define complexity purely in terms of objective task qualities such as the number of subtasks (March and Simon 1958), the number of roles that need to be satisfied (Campbell 1984), and the time taken to perform a task (Lopata et al. 1985). The Department of Labor’s Dictionary of Occupational Titles and O*NET also measure the interpersonal, operational, and problem-solving dimensions of work complexity, which vary widely across jobs (Kohn and Schooler 1973, 1978). Irrespective of how complexity is measured, however, scholars agree that complexity is a job characteristic that makes work meaningful.
In our setting, we observed that complexity was understood as the time taken to perform a specific garment task: simple operations took only seconds to execute, in contrast to complex operations, which could take several minutes. In line with the predictions from existing literature, our interviews with workers engaged in simple versus complex operations revealed variation in how meaningful they understood their work as being. For example, interviews with workers engaged in various simple operations revealed that they found their job to be “boring,” “repetitive,” and “monotonous,” whereas in other interviews, workers performing more complex operations described their job as “fulfilling” and “challenging.”

Our observation further revealed that work complexity appeared to govern how workers responded to the monitoring intervention. For example, one worker performing a simple “fly blind hem” operation said, “My work was boring before . . . I like pressing the various buttons [on the machine] and this makes the work fun.” Another worker engaged in simple trimming said, “Today if I do 80%, then tomorrow I try for 81% . . . in this way, each day is a competition!” It seemed like workers engaged in simple operations found their work to be monotonous and as such, monitoring through its sophisticated technology and feedback mechanism allowed these workers to gamify their work. Moreover, this lens of seeing monitoring as a game seemed to intrinsically motivate workers performing simple work to be more productive. A worker engaged in basic “assembly loading” said, “The machine has helped me realize my potential . . . I am delivering better production numbers than I ever thought I could do!”

Researchers have also noted that workers often engage in “job crafting,” whereby they utilize opportunities to alter how they perceive their tasks to foster engagement with their work (Berg et al. 2010, Leana et al. 2009). One type of job crafting that has been studied is called “gamification,” whereby workers transform work tasks into play to make the affective experience of work more fun (Deterding et al. 2011). The idea is that workers gamify work that they perceive as unmeaningful or monotonous, and that this gamification often motivates better performance (Burawoy 1979, Roy 1952, 1959, Sherman 2007). The intrinsic motivational power of gamification has even led managers
to introduce gamification into the workplace by incorporating design elements from games into productive work processes (Deterding et al. 2011, Edery and Mollick 2009, Mollick 2014). Gamification has further been shown to promote learning and improve performance because of workers’ desire to improve their skills and receive positive feedback (e.g. Hackman and Oldham 1976, Burgers et al. 2015). Therefore, based on our field observation along with these insights from existing literature, we offer the hypothesis:

**HYPOTHESIS 1.** The effect of monitoring on worker productivity will be positive in the context of simple work.

In contrast, our observation indicated that workers engaged in complex work responded very differently to monitoring. These workers took pride in their work and valued having autonomy in performing their relatively more time-intensive operations. For example, one worker sewing intricate internal pockets said, “Operators like me deliver because of our deep familiarity and interest in our operation.” As such, workers engaged in complex work viewed monitoring as unnecessarily controlling and interfering with their rhythm and their preexisting relationship with their work. One worker performing a complex “lapel seam” operation said, “My work is critical and difficult to do . . . having industrial engineers observing while work is being done would make even a normally fast operator slow down.” Another such worker said, “Before the machine was introduced, I was able to allocate my time on my own, . . . but now the countdown has added additional work pressure.” Further, this interpretation of monitoring as micromanagement seemed to crowd out workers’ intrinsic motivation, resulting in lower performance. One worker performing the “under collar zig zag stitch” said, “I do not like having the watchman machine . . . I used to really like my operation, now I’m not so sure . . . [I] think that it has not positively impacted my production levels.”

Some research has also suggested that when work is nonstandard, monitoring can be construed as micromanagement and can thereby crowd out intrinsic motivation (Manso 2011, Alvesson and Sveningsson 2003, Ederer and Manso 2013, Frey and Jegen 2001). This
research has further shown that monitoring can, in this way, lead to lower productivity (Azoulay et al. 2011). Another strand of research has argued that monitoring may also lead workers engaged in meaningful work to question their competence, detracting from their performance (Lawler III 1992). This research has described how micromanagement may impinge on workers’ autonomy and, through this channel, impair workers’ intrinsic motivation and job performance (Spreitzer 1996, Deci and Ryan 1975). Thus, we offer the hypothesis:

Hypothesis 2. The effect of monitoring on worker productivity will be negative in the context of complex work.

To test these hypotheses, we measure the effect of monitoring on productivity in lines and jobs of varying complexity. Moreover, the setting allows us to focus narrowly on monitoring and rule out some important mechanisms, especially those emphasized in the economics and management literatures that tie monitoring to rewards or punishments. Although the setting and intervention are less able to finely distinguish among the many narrower psychological mechanisms explored in the lab that could be consistent with the qualitative data presented above, the study does suggest that certain mechanisms may be more applicable to this case than others. We return to this discussion after presenting the results.

3. Quantitative Data and Methods

3.1. Data on Line and Operational Productivity

In this setting, we obtained two sets of productivity data: line-level data and operation-level data. Line-level productivity data include daily productivity at each of the twelve lines from 2009 to 2014. Operation-level productivity data include daily operation-level productivity within pant lines for a subset of dates in 2012. We exclude jacket lines in the operation-level analysis because we do not have pretreatment productivity data for jacket production. Fortunately, for the pant lines, we have some operation-level data both before and after the treatment. These data were collected
through a manual tracking system conducted by nonsupervisory staff; this system predated the treatment and continued, without changes in how the data were collected, in the posttreatment period. We use these manual data to exclude the possibility that our estimated treatment effect is biased by the data collection regime. The empirical strategy will exploit variation in complexity across product lines and within the pant lines. Other standard HR data are available for both the pretreatment and posttreatment periods.

Within each line, production is executed through a set of operations of varying complexity. As is common among garment manufacturers, our plant rates the complexity of operations according to their standard minute values (SMVs), a standard unit for measuring the complexity of an operation. Specifically, an SMV represents the number of minutes allowed for a qualified worker working at a standard performance level to complete a given operation, with higher values corresponding to more complex operations requiring greater time. Operations rated with higher SMVs typically require a greater number of subtasks, greater dexterity, and greater skill to complete. As is well-known in the garment manufacturing industry, complexity (and therefore rated SMVs) also varies by the type of garment, with the average operation on men’s jacket lines rated as more complex than those on pant lines. Similarly, in this plant too, the SMVs of jacket operations have a mean of 0.74 and standard deviation of 0.36, whereas the SMVs of pants operations have a mean of 0.59 and a standard deviation of 0.31. Figure 3 illustrates the variation in work complexity across and within product lines.

Productivity is measured by “percent efficiency,” which depends on both the speed of the operator and the complexity of the task. For example, if the manufacturer rates an operation as having an SMV of 0.75, then it expects a skilled operator to complete one operation in 0.75 minutes. If a particular worker takes an average of 0.50 minutes to complete an operation over a shift, then the percent efficiency for that worker is 150%.

---

3 SMVs are also referred to as standard allowed minutes, or SAMs.
At the end of the production line, inspectors check all products and return any with deficient quality to the responsible worker. This rework reduces a worker’s measured efficiency, in that rework takes time but does not count toward items produced. The SMV benchmark also includes expected rework time, along with rest and other allowances. Because the percent efficiency includes both initial production and rework, plant management treats efficiency as a measure of both quantity and quality.

Prior to the treatment period, daily productivity in all lines was monitored at the line level. This was done by comparing the number of fully finished garments produced by a line to the daily target for that line. This system of tracking line productivity continued unchanged posttreatment as well. These records provided the pre- and posttreatment line-level productivity data.

In the pretreatment period, supervisors could informally monitor individual workers. More systematic monitoring at the individual level within lines was done at random intervals with the aid of a manual system in which a data entry clerk (who had no formal authority) asked workers to report how many items they produced. These manual records provided the pretreatment operation-level productivity measure for the pant lines.

Importantly, in the treated lines post-treatment, the plant supervisors continued to use the manual system for process control, despite the new RFID technology and data, for three main reasons. First, supervisors split their time across lines, and the new monitoring technology was only available for treated lines. Second, supervisors simply did not have easy access to the data; they spent their time on the shop floor, but the RFID data were only accessible through a computer on a separate floor. Third, there was inertia to move away from the old process control system. As a result of this, the manual records provided the posttreatment operation-level data as well.

In fact, from 2009 to 2014, among the 2,212 workers, there was no evidence that any process control data were cited in any formal disciplinary action. In this entire period, there was only one involuntary termination, which cited a behavioral issue unrelated to job performance. For these reasons, it appears that the most obvious treatment effect occurred
because of workers’ interaction with the monitoring system rather than through direct retribution from management.

3.2. Empirical Strategy: Difference-in-Differences

The analysis consists of three parts: (1) the unconditional analysis, where we examine overall changes in productivity at treated lines posttreatment; (2) the analysis that conditions on the complexity of the product line; and (3) the analysis that conditions on the complexity of worker-level operations within the pant lines.

First, we estimate the effect of the monitoring intervention on productivity, before conditioning on complexity:

$$Y_{it} = \beta_0 + \beta_1 \text{post}_{it} + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \phi_i + \tau_t + \epsilon_{it}$$

(1)

In equation (1), $Y_{it}$ is the productivity (the percent efficiency) of line $i$ on day $t$, for $i$ in 1 through 12 and $t$ in 1 through 1871; treated$_i$ is an indicator that takes a value of 1 if treated$_i$ is ever a treated line; and post$_{it}$ is an indicator for whether lines producing the same product (jackets or pants) as line $i$ were in the posttreatment period, regardless of whether that line was actually treated. Because of the technology’s staggered introduction, this variable takes a value of 1 for lines producing pants after October 1, 2012 (for $t > 1126$), or for lines producing jackets after December 1, 2012 (for $t > 1174$). Then, posttreated$_{it}$ is an indicator that takes a value of 1 if treated line $i$ is in the treatment period at time $t$, which is functionally equivalent to the interaction between the posttreatment and treated line indicators. We run these three regressions with and without line, month, and day of week fixed effects, respectively denoted by $\phi_i$ and $\tau_t$.

We cluster standard errors by line to allow for errors to be correlated within lines.

As in a standard experimental difference-in-differences framework, the identification comes from the difference between treated and nontreated lines before and after the treatment. This yields estimates for the causal effect of the treatment in the case that,

4 Note that $t$ denotes days, but we estimate fixed effects at the month level.
whatever the *ex ante* differences between the lines, the treatment is not correlated with other factors also affecting productivity. Although the treated lines in this case were not chosen at random, they were chosen for a reason unrelated to their productivity: the treated lines are closest to the technical support offices, which aided technology installation and maintenance. Because technical support and line supervisors are in different departments, and because lines are otherwise homogeneous, it is unlikely that this factor is also correlated with productivity.\(^5\)

Likewise, Table 1 shows that treated and nontreated lines were observationally similar in terms of their worker characteristics.

Next, we examine whether the effect of monitoring depends on work complexity. For this, we rerun equation (1) separately for lines producing jackets and pants:

\[
Y_{pt} = \beta_0 + \beta_1 \text{treated}_p + \beta_2 \text{post}_p + \beta_3 \text{posttreatment}_p + \phi_p + \tau_p + \epsilon_{pt} \quad (2a)
\]

\[
Y_{jt} = \beta_0 + \beta_1 \text{treated}_j + \beta_2 \text{post}_j + \beta_3 \text{posttreatment}_j + \phi_j + \tau_j + \epsilon_{jt} \quad (2b)
\]

where \(p\) denotes the nine pant lines in which workers perform relatively simple operations, \(j\) denotes the three jackets lines in which workers perform relatively complex operations, and all other variables remain the same. Hypothesis 1 concerns the direction of the coefficient on \(\beta_3\) in the first equation, and Hypothesis 2 concerns the direction of the coefficient on \(\beta_3\) in the second equation.

Equations (2a) and (2b) are convenient for comparing the productivity growth of treated lines versus nontreated lines, separately for the lines performing simple and complex operations. However, we may also wish to test whether the effect of the monitoring technology is different between lines performing simple versus complex operations. For this, we fully interact equation (1) with an indicator that the line is producing jackets:

\[
Y_{it} = \beta_0 + \beta_1 \text{post}_i + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_i + \phi_i + \tau_i + \epsilon_{it} + \beta_5 \text{complex}_i + \beta_6 \text{posttreated} \ast \text{complex}_i + \beta_7 \text{treated} \ast \text{complex}_i + \beta_8 \text{posttreated} \ast \text{complex}_i + \phi_i + \tau_i + \epsilon_{it} \quad (2c)
\]

\(^5\)More importantly, the difference-in-differences test assumptions would be violated only if proximity to technical support affected productivity and this effect changed over time.
where complex, denotes that line \( i \) is producing the more complex product (jackets), and all other variables remain as before. As a three-way interaction, remember that \( \beta_0 \) through \( \beta_3 \) are estimated with the noncomplex (pant) lines. As before, we run these three regressions with and without line, month, and day of week fixed effects. We cluster standard errors by line.

This line-level analysis allows us to examine how monitoring affected productivity depending on the average complexity of the tasks performed within the lines. One threat to validity of this analysis is that we do not have an infinite number of lines and treatments; if lines possess their own idiosyncratic time trends, then we may observe an effect by chance. To address this, we cluster standard errors by line, (1) additionally adopting the relatively conservative \( T(G - L) \) critical value thresholds, (2) explore results allowing for errors to be correlated across lines, not just within lines, with a wild cluster bootstrap, and (3) perform a complete set of placebo tests using all counterfactual combinations of two nontreated lines.

Our operation-level analysis provides an even stronger check to this issue: we examine how monitoring affected productivity depending on the complexity of the individual operations within the pant lines, which moves beyond comparing pant and jacket lines. These operation-level results are not as vulnerable to the problem of estimating effects with few clusters. In particular, we exploit variation in operational complexity within pant lines. For example, attaching pant waistbands is a relatively complex operation with an SMV of 0.95, meaning an operator must execute an operation every 57 seconds (including rework and rest) to achieve 100% efficiency. In contrast, attaching pant zippers is a relatively simple operation with an SMV of 0.74, corresponding to performing the operation in 44.4 seconds. The boxplot in Figure 3 presents the distribution of work complexity within pant lines, along with other examples of

---

6 Specifically, we adopt critical values from the \( T(G - L) \) distribution, where \( G \) is the number of clusters and \( L \) is the number of time-invariant regressors, specifically the treatment and product indicators where they apply. This yields more conservative thresholds for significance than the \( T(G - 1) \) distribution. For a discussion of these methods, see Bertrand et al. (2004), Cameron et al. (2008), Cameron and Miller (2015).
operations. We are interested in whether the change in efficiency on treated lines depends on the complexity of the operation.

For these tests, our empirical strategy mirrors what we use for line complexity. Specifically, we begin by splitting pant operations into simple operations and complex operations, depending on whether they are respectively below or above the 75th percentile SMV among pant operations. We run the regression using the 75th percentile cutoff (corresponding to an SMV of 0.75) because pants operations are already relatively simple, so we denote complex pant operations to be the most complex among them.

The regressions take the form:

\[ Y_{ist} = \beta_0 + \beta_1 \text{treated}_i + \beta_2 \text{post}_ist + \beta_3 \text{posttreatment}_ist + \phi_i + \tau_t + \epsilon_{ist} \]  

(3a)

\[ Y_{ict} = \beta_0 + \beta_1 \text{treated}_i + \beta_2 \text{post}_ict + \beta_3 \text{posttreatment}_ict + \phi_i + \tau_t + \epsilon_{ict} \]  

(3b)

where \( Y_{ist} \) and \( Y_{ict} \) represents the efficiency for line \( i \)'s simple operation \( s \) or complex operation \( c \) at month \( t \), and all other variables are as they were before. We run these regressions with and without line, month, and day of week fixed effects and standard errors clustered by operation.

Likewise, we perform a fully interacted regression:

\[ Y_{iott} = \beta_0 + \beta_1 \text{post}_it + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_it + \beta_4 \text{complex}_{io} + \beta_5 \text{post}_it \times \text{complex}_{io} + \beta_6 \text{treated}_i \times \text{complex}_{io} + \beta_7 \text{posttreated}_it \times \text{complex}_{io} + \phi_i + \tau_t + \epsilon_{pt} \]  

(3c)

where \( Y_{iott} \) is the percent efficiency for line \( i \)'s operation \( o \) at month \( t \). For consistency and completeness, we show results where complex\( o \) is treated as a discrete variable indicating whether the SMV rating of the operation is above the 75th percentile SMV among pant jobs, and additionally, where complex\( o \) is a continuous variable representing the SMV rating of the operation.

This operation-level analysis has three distinct advantages over the line-level analysis. First, pant and jacket lines could be correlated with some factor other than
average complexity that yields productivity gains or losses from monitoring; switching to
the operation-level analysis allows us to evaluate complexity directly. Second, there are a
large number of operations, and therefore a relatively larger number of clusters in the
operation as compared to line analysis, which affords much greater statistical power.
Third, since our measure of complexity is a continuous variable, the operation-level
analysis allows us to estimate the effect of monitoring at discrete intervals.

4. Results

4.1. Complexity across Lines

We begin by showing the mean efficiency at treated and nontreated lines before and
after the monitoring technology was implemented. We do this for all lines, by product
line, and by operational complexity within pant lines. We also present estimates for the
first-differences and difference-in-differences, along with the standard errors treating the
line-day as the unit of analysis. For these summary statistics, we do not cluster standard
errors.

[Table 2]

The first numbered panel of Table 2 shows that overall productivity improved at both
treated and nontreated lines, although the overall improvement was slightly greater at
treated lines. Note that the mean percent efficiency is not 100% but rather around 60%. As
such, a gain of 4% efficiency generally corresponds to about a 7% gain in productivity (64%
/ 60% = 107%). Also note that the overall differences reported in this first panel include
both the direct effect of monitoring on pant and jacket lines and also the compositional
effect that places greater weight on pants, which compose nine of the twelve lines.

The second panel (2a) shows changes in productivity among the two treated pant lines
and seven control pant lines. Productivity remained relatively flat at the nontreated lines
and substantially rose at the two treated lines after the introduction of the monitoring
technology. The next panel (2b) shows changes in productivity among the treated jacket
line and two control jacket lines. In contrast to the pant lines, productivity at the treated jacket lines declined even as productivity at the nonteated lines improved.

The third panel shows changes in productivity within the 51 pant operations. Efficiency for operations with complexity below the 75th percentile complexity (in 3a) remained approximately constant within the nonteated lines. However, efficiency at treated lines rose by 5.1% posttreatment. For complex operations (in 3b), efficiency slightly declined at nonteated lines and declined by a larger amount at treated lines.

Next, we turn to the unconditional regression results before conditioning on line or operational complexity. Note that regression estimates depart from the means presented earlier because the full regression results further control for line, month, and day of week effects, and the regression results cluster standard errors at the line level to allow for correlated observations within lines. These regressions correspond to equation (1) in the empirical strategy section. Table 3 presents the results.

| Table 3 |

Table 3, column 1, shows the classic difference-in-differences without line or month controls. Treated lines had slightly higher productivity prior to the implementation of the monitoring technology, but no statistically significant growth after the technology was introduced. Column 2 introduces month and day of week fixed effects. Note that month fixed effects are collinear with the post-period and so the interpretation remains similar except variance is absorbed across 72 months, rather than two periods. Column 3 introduces line fixed effects, which are collinear with being a treated line. Similarly, the interaction term here should be interpreted as the difference in productivity among the three treated lines after the treatment takes effect. Column 4 introduces line, month and day of week fixed effects. Given that we have a balanced sample and the treatment is largely uncorrelated with the controls, it is not surprising that the main coefficient of interest, the difference-in-differences estimate, changes little as we introduce finer controls. This estimate is small and not statistically different from zero.
Stopping here, we might conclude that the monitoring technology had no effect on productivity. However, the following tests examine the heterogeneity of the treatment effect by product and operational complexity. We begin the conditional analysis by examining the effect of monitoring on the simple product (pants) and the more complex product (jackets). These correspond to equations 2a - 2c in the empirical strategy section. Table 4 presents results.

Table 4, columns 1 and 2 correspond to Equation 2A in the empirical strategy section and show the results of the difference-in-differences analysis on the pant lines. Efficiency at treated pant lines rose an estimated 8.06% relative to the nontreated lines. Given that these lines were operating at 64.2% efficiency before the treatment, relative productivity improved by about 13.7%. The improvement is also statistically significant. Under the conservative specification, with clustered standard errors by line, standard errors are less than 3%. Results for pant lines are substantively similar after including line, month, and day of week fixed effects.

Table 4, columns 3 and 4 correspond to equation 2b in the empirical strategy section, and show results restricting analysis to the jacket lines. Note that we now have one-third of the full sample observations because only one-third of the lines produce jackets. Compared to nontreated jacket lines, productivity at the treated jacket line dropped by 9.4%. This represents a reduction of about 13.7% had the treated lines otherwise followed the same efficiency trajectory as the nontreated lines. Again, specifications are similar if we include line, month, and day of week effects. Taken together, columns 1 through 4 suggest that the effect of the treatment was significantly different from zero at both lines performing simple and complex operations, albeit in opposite directions.

Table 4, columns 5 and 6, correspond to equation 2c in the empirical strategy section and use triple-differences to test whether the treatment effect was significantly different, depending on whether the line performed simple or complex operations. Before we interpret the main result, a few characteristics deserve note. First, mechanically, the triple-difference
coefficient will be approximately equal to the difference between the two-way interactions in the prior columns. Second, we can estimate $Post \times Treated$ and $Post \times Treated \times Complex$ in the final column because of the staggered treatment. Third, in even numbered columns, the treated and post terms are respectively absorbed by the line and month fixed effects. Now turning to the coefficient of interest, the triple-differences model estimates that the difference in the effect of monitoring between lines performing simple and complex operations is 17.5% in the reduced model and 15.7% after including the fixed effects. Both are statistically different from zero.\footnote{We also perform these tests with the wild cluster bootstrap-t procedure. While our clustering procedure allows for errors to be correlated within clusters (lines), it assumes that errors are independent across clusters. In the full model, errors may be correlated across lines as long as they are absorbed by the time effects. The wild bootstrap is a more conservative test that further allows for correlated errors across clusters and retains the properties of asymptotic validity and refinement, thereby reducing the likelihood of yielding a false positive (Liu et al. 1988, Mammen 1993). With the wild cluster bootstrap, in model 5, the coefficient on $Post \times Treated$ is statistically significant with $p < 0.01$, and the coefficient on $Post \times Treated \times Complex$ is significant with $p < 0.05$, but no longer with $p < 0.01$. In model 6, these statistics are respectively $p < 0.05$ and $p < 0.01$. When the bootstraps impose the null, as advocated by Cameron et al. (2008) and Davidson and MacKinnon (1999), these statistics in models 5 and 6 are respectively $p < 0.01$, $p < 0.10$, $p < 0.01$, and $p < 0.05$. Overall, we conclude that results are robust to the decision to allow residuals to be correlated between clusters.}

Figure 4 illustrates the weekly difference in efficiency between treated and nontreated lines. As shown in the top panel, the two treated pant lines were generally more productive than the nontreated lines even prior to the treatment. However, the productivity gap grew strongly after the treatment went into effect. The pretreatment difference among jackets is less pronounced, and the productivity gap turned negative after the treatment, thus mimicking the regression results.

4.2. Complexity within Lines

Next, using operation-level data from the pant lines, we advance our analysis by examining how the effect of the monitoring technology varies by operational complexity. These correspond to Equations 3A-3C in the empirical strategy section. Table 5 presents results.
Table 5, columns 1 and 2, restrict analysis to the relatively simple pant operations: operations with SMV values less than or equal to 0.75, which correspond to the 75th percentile among pant operations. Once again, the post and treatment terms drop out in Column 2 because they are absorbed by line, month, and day of week fixed effects. Given that the typical pant operation is simpler than the typical jacket operation, these operations largely rank among the most simple operations performed at the plant. Compared to nontreated pant lines, efficiency rose at treated lines posttreatment. The coefficients in columns 1 and 2 are moderately significant ($p < 0.1$) and substantively large (the estimated increase in efficiency of 4.97% and 5.36% is about an 8.2% and 8.9% gain relative to the prior efficiency among simple operations with 59% mean efficiency).

Table 5, columns 3 and 4, restrict analysis to pant operations with SMV values above the 75th percentile among operations at the pant lines (SMV > 0.75). Compared to nontreated pant lines, efficiency declined by 4.15% and 3.58%, before and after including detailed fixed effects. These differences are respectively significant with $p < 0.05$ and $p < 0.01$, and are substantive, representing declines in efficiency of 6.2% and 5.4% relative to the prior efficiency among complex operations.

Table 5, columns 5 and 6, include all operations and test whether the coefficients are different between simple and complex operations. Mechanically, the coefficient of interest, the triple-difference, is the sum of the two coefficients of interest in the prior models (i.e., $-4.15 - 4.97 = -9.13$, with rounding). The resulting coefficients are substantial (respectively representing differences of more than 15% and 10% of mean productivity at pant lines) and statistically significant at 1% in the reduced model and 10% in the full model.

---

8 Again, we can allow for differences in variance across clusters using the wild cluster bootstrap-t procedure. In model 5, the coefficient on $Post \times Treated$ is statistically significant with $p < 0.01$, and the coefficient on $Post \times Treated \times Complex$ is significant with $p < 0.10$. In model 6, these statistics are $p < 0.01$ and $p < 0.05$. When the bootstraps impose the null, the statistics for models 5 and 6 are $p < 0.01, p < 0.10, p < 0.10$, and $p < 0.05$. 
Whereas the product-level analysis features only two products, the operation-level analysis features 51 pant operations with varying levels of SMV ratings. In principle, this allows us to estimate the effect of the monitoring technology at different intervals of complexity. In practice, this is difficult because we are now relying on a thin slice of the data for each treatment effect: workers performing operations within a narrow SMV interval, at treated lines, posttreatment. To do so, we continue to use the most conservative assumptions: we estimate the difference in percent efficiency at treated lines posttreatment at eight discrete SMV values, rounding the SMV values of each operation to the nearest tenth. Functionally, this is equivalent to running the regression in Table 5, column 6, except with interaction terms for each discrete bin of complexity at 0.1 SMV interval. We present coefficient estimates and standard errors in Figure 5.

[Figure 5]

As shown in Figure 5, the error bars are relatively large within any one category of SMV values. Nonetheless, results suggest there may be a negative relationship between complexity and the effect of the monitoring technology on productivity. Consistent with the earlier regression results, this is primarily driven by gains in productivity among the most simple operations.

4.3. Line Placebo Analysis

For the line-level analysis, in a single-differences framework, identification would come: (1) from posttreatment productivity minus pretreatment productivity within treated lines, or (2) from treated line productivity minus control line productivity within the posttreatment period. These strategies would respectively require time effects to be capture by pretreatment trends, or for treated lines to be otherwise identical with respect to factors affecting productivity. The strength of the difference-in-differences strategy is that it allows us to use control lines to distinguish time effects and the two time periods to distinguish line effects. Rather, the threat to validity in difference-in-differences is that
treated and control lines would follow different time trends in the absence of the treatment. This requires the parallel trends assumption.

Our parallel trends assumption relies on two important features of this setting: lines were technologically identical prior to the treatment, and the rule governing which lines were treated appears to be uncorrelated with any reasons related to productivity or line characteristics that may affect future productivity. This serves as a check against the usual concern with difference-in-differences: that the intervention is endogenous. However, we also have a more subtle concern. In an experimental framework with large samples, we would satisfy the parallel trends assumption through random assignment and the law of large numbers. However, given our smaller sample of lines, we may be worried that these lines also follow their own idiosyncratic time trends that would violate parallel assumptions despite the orthogonal assignment to the treatment. These issues are discussed in Bertrand et al. (2004).

Up to this point, we’ve allowed for idiosyncratic time trends by line using higher critical values for our significance tests, through the wild-cluster bootstrap, and through the operation-level analysis. Following Bertrand et al. (2004), we also check our results against line placebo interventions. Our procedure goes as follows. There are nine pant lines, two of which are truly treated; we first indicate these two truly treated lines as being untreated. Then, from the remaining truly untreated seven lines, we sequentially indicate every combination of two lines as treated lines, giving us a total of 21 placebo treatments (7 choose 2). Lastly, we see how often this procedure reproduces the results observed when we properly indicate the truly treated lines. We do the same for the jacket lines. Note that there are 3 jacket lines, one of which is treated; so the same procedure yields 2 placebo treatments for the jacket lines. Since the placebo interventions are uncorrelated with the true intervention, we should find our same result no more than we would expect by statistical chance. Alternatively, idiosyncratic line-specific time trends, coupled with a finite sample of treated lines, would generate a large number of false positives.

---

9 We do not have the corresponding issue with the operation-level analysis since operational complexity is a continuous variable and our analysis evaluated the effect of complexity at semiparametric intervals.
positives among the placebos. In Figure 6, we rank order the placebo and nonplacebo coefficients for the Post × Treated term and present these with their 95% confidence intervals.

[Figure 6]

For the pant lines, none of the 21 placebo treatments generates a positive and statistically significant result with \( p < 0.05 \), although one placebo comes close (\( p = 0.053 \)). This placebo treatment features the two nontreated lines that exhibited the largest growth in productivity during the treatment period, not counting the two treatment lines. Two of the 21 combinations of placebo treatments yield a statistically significant decline relative to the other lines. For the jacket lines, the effects of the placebo treatments are not significantly different from zero. Overall, we conclude that the line placebos do not yield false positives at a greater rate than we would expect by statistical chance.

5. Discussion

Our fieldwork informed our hypotheses regarding how the effect of monitoring would depend on the complexity of work. Consistent with these hypotheses, we found that productivity improved among lines and jobs performing simple work and declined among lines and jobs performing complex work. We conclude that this instance provides clear field evidence that depending on key job characteristics such as complexity, the act of monitoring itself can improve, reduce, or have no net effect on productivity.

We then ask: why did the effect of monitoring on productivity vary by complexity? We begin with what our study largely rules out. For our entire study period, 2009-2014, at no point was monitoring or any performance information used for promotions, incentive pay, terminations, or other conditions of employment, which allow us to rule out a broad swath of well-evidenced theoretical reasons to expect monitoring will affect productivity. Indeed, prior theoretical work on shirking (e.g. Shapiro and Stiglitz 1984), multitasking (Holmstrom and Milgrom 1991), complementary performance management practices
(Brynjolfsson and Hitt 2000), and monitoring as a deterrent for experimentation and innovation (Bernstein 2012, Campbell et al. 2011) all rely on the assumption that monitoring may be used for reward or punishment. In other words, our results imply that theory need not invoke reward or punishment to explain changes in productivity following a monitoring intervention. The act of monitoring itself is indeed sufficient to affect productivity.

But what mechanisms does our study rule in? Our fieldwork suggests that monitoring enabled the gamification of simple work and crowded out workers’ intrinsic motivation in the case of complex work. Interviews with workers suggest that part of this effect may owe to a special circumstance of the monitoring intervention: it allowed workers to see their productivity in real time, which had varying motivational effects depending on whether work was complex. Indeed, in line with our qualitative data, Deci and Ryan (1975) and self-determination theory have found that monitoring can affect workers’ intrinsic motivation, a result that has been widely replicated in the lab.

Our results may also be contextualized within the broader sociopsychological literature on monitoring, which largely draws from laboratory studies. Social facilitation theory (Zajonc et al. 1965) proposes that workers have low levels of arousal when performing routine work, and therefore the presence of an audience (e.g. a supervisor) raises arousal, and there is an inverted-U relationship between arousal and performance. As such, productivity is locally increasing with respect to stimulation (and monitoring) for routine work, until the point at which it begins to decline. Unfortunately, testing this specific mechanism (and related mechanisms, such as cognitive overload) would require measures of stress and arousal, which we do not have. Nonetheless, our study brings new and unique evidence for evaluating the differential effect of monitoring on productivity by work complexity in relation to these extant laboratory studies. First, prior studies have almost exclusively relied on human rather than computer monitors. Second, prior studies have also spanned relatively little time, for example, testing the effect of monitoring over the span of an hour. It is less clear whether test subjects or workers would remain aroused over a long period of time or would alternatively become acclimated to a human
(or computer) monitor. Our study features a computer monitor over a long time in the field and can therefore speak to the long-term consequences of electronically monitoring workers, thus advancing the existing literature. Further, as a field rather than a lab study, external validity of our findings is also less of a concern.

Another theory suggests that monitoring will improve, not reduce, productivity among more complex tasks. Closer monitoring may allow supervisors to assist workers when problems arise, which may be more likely when work is more complex [Bell and Kozlowski 2002, Garicano 2000]. Bonet and Salvador (2017) reported evidence for this theory in a field setting, where IT service technicians suffered greater declines in productivity the more distant their job assignments were, especially as the complexity of their tasks rose. However, this theory primarily concerns itself with physical, in-person monitoring where workers see supervisors as problem solvers and is therefore less applicable to our context where workers experience monitoring through technology without the accompanying problem-solving benefits.

Finally, we turn to another question raised in our analysis: are differences in the results being driven by complexity or by something else correlated with complexity? If we could only perform the line analysis, we may be concerned that our results are driven by other differences between pants and jackets besides complexity; fortunately, we can replicate our results from the line analysis with individual-level analysis within pants production using a large number of tasks. Therefore, any alternative explanation should explain why our results are robust not only across pant and jacket lines, but also across operations within the pant lines. Further, given our precise measure of complexity in this setting—the time required to complete an operation—if something other than complexity is driving our results, it should also be correlated with the time expected to complete tasks.

5.1. Contributions

In general, the decision to monitor speaks to many aspects of the employment relationship: economists have highlighted how monitoring allows firms to tie performance
measures to rewards and punishments (Hölmstrom and Milgrom 1991) and sociologists have highlighted how monitoring can affect workers’ identity and sense of professional autonomy (Burawoy 1979, Roy 1959). Not surprisingly, then, monitoring has long played a central role in management research. In fact, as technological constraints on monitoring relax, employers must increasingly ask themselves not whether they can monitor workers but whether they should. In light of this, our paper makes four important contributions.

First, we shift the debate away from the question of whether monitoring affects worker productivity and toward determining the conditions under which monitoring increases or decreases productivity. Recent studies have largely affirmed intensive monitoring and measurement as a “good” management practice (e.g. Pierce et al. 2015, Staats et al. 2017). Even so, some work has suggested that monitoring could also be detrimental to productivity, especially in isolation of other practices (e.g. Bernstein 2017). Our paper helps bridge the two perspectives in the literature by demonstrating that depending on key characteristics of the job being performed, the same monitoring technology can sometimes be effective in increasing worker productivity but at other times, be counterproductive. In this way, our paper calls for the literature to move into investigating the conditions moderating the effect of monitoring on productivity.

Second, existing studies on monitoring and productivity have been conducted in diverse settings, ranging from restaurants (Pierce et al. 2015) to high-tech manufacturing (Bernstein 2012), and these studies have found divergent results pertaining to the effect of monitoring on worker productivity. Our paper suggests that increased attention to the work being monitored might help explain the divergent results in the literature. In particular, our paper uncovers the importance of paying attention to job characteristics such as work complexity and demonstrates how the effect of monitoring varies dramatically by work complexity. Exploiting variation in work complexity within garment production, we find that monitoring improves productivity for simple work, but hinders productivity for complex work. Our review of the literature reveals that across different studies too, there is wide variation in the complexity of the work being
monitored, and further that the negative effect of monitoring on productivity has been emphasized in studies investigating relatively more complex work (e.g., Bernstein [2012]). Thus, focusing on job features like work complexity can help explain the divergent effects of monitoring on productivity documented in the literature. Further, the moderating effect of work complexity may be even more salient in other contexts because both the simple and complex garment operations in our study are arguably less complex than typical jobs from the universe of settings in which we may wish to study the effect of monitoring on productivity.

Third, this study contributes to the monitoring literature by demonstrating that we do not need to invoke rewards or punishments to explain why monitoring can positively or negatively affect productivity. This implies that a wide range of incentive-based monitoring theories—such as those involving pay for performance, terminations for shirking, or punishments for productive rule-breaking—may be true but are nonetheless unnecessary. Instead, our results imply that we must invoke some other mechanism to explain the effect of monitoring on productivity, such as changes in workers’ intrinsic motivation. Indeed, Deci and Ryan (1975) and self-determination theory have posited that monitoring can affect workers’ intrinsic motivation; we push this idea even further by positing that monitoring might differentially affect workers’ intrinsic motivation depending on the complexity of the work being performed or other such job characteristics that make work meaningful.

Finally, we also offer a methodological contribution for audiences interested in the Hawthorne effect. Of course, the Hawthorne effect has been of great interest to researchers, since we typically rely on interventions that include monitoring as part of data collection. And yet, whether the Hawthorne effect exists has been the subject of debate (see, for example, Adair [1984] Levitt and List [2011]), though empirical work continues to consider it. We present evidence that checks against the Hawthorne effect remain warranted. However, as our case points out, the effect of monitoring may be positive, negative, or neutral, depending on the degree of work complexity.
References


Mayo E (1933) The Human Problems of an Industrial Civilization (New York, Macmillan Co.).


Strauss AL, Corbin J (1990) *Basics of Qualitative Research* (Sage).


Taylor FW (1914) *The principles of scientific management* (Harper).

Turner AN, Lawrence PR (1965) *Industrial Jobs and the Worker* (Harvard Univ.).


Zajonc RB, et al. (1965) *Social facilitation* (Research Center for Group Dynamics, Institute for Social Research, University of Michigan Ann Arbor).
Figure 1  Study Design

Jan 1, 2009

<table>
<thead>
<tr>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated Lines (n=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Lines (n=7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated Lines (n=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Lines (n=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pants (simple)

Jackets (complex)

Data begin

Monitoring technology implemented on treated jacket lines

Monitoring technology implemented on treated pant lines

Oct 1, Dec 1, 2012 2012

Dec 31, 2014

Data conclude
Figure 2  Technology to Monitor Workers: RFID Scanners Installed on Treated Lines
Figure 3  Boxplot showing work complexity of operations across and within lines

<table>
<thead>
<tr>
<th>SMV</th>
<th>Jackets (complex)</th>
<th>Pants (simple)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Lining to arm hole (1.49)</td>
<td>Loop bartack (1.11)</td>
</tr>
<tr>
<td>b</td>
<td>Facing seam (0.89)</td>
<td>Front panel overlock (0.71)</td>
</tr>
<tr>
<td>c</td>
<td>Trim lapel corner (0.71)</td>
<td>Fix hook (0.55)</td>
</tr>
<tr>
<td>d</td>
<td>Front dart mark (0.45)</td>
<td>Back button hole (0.41)</td>
</tr>
<tr>
<td>e</td>
<td>Under collar gathering (0.19)</td>
<td>Wash care label (0.19)</td>
</tr>
</tbody>
</table>

Note. The boxplots represent the distribution of work complexity (SMVs) in jacket lines and pant lines, including tail operations and quartiles.
Figure 4  Difference in Line Productivity between Treated and Nontreated Lines (Weekly)

Note. Vertical lines indicate the treatment period. Horizontal, dotted lines represent mean differences in line productivity pre- and post-treatment.
Figure 5  Difference-in-Differences Estimates for Effect of Monitoring on Productivity in Pant Lines by Operation Complexity Bins

Note. The figure presents difference-in-differences coefficient estimates and standard errors for the effect of treatment on efficiency, by operational complexity (rounded to the nearest 0.1 SMV).
Figure 6  Placebo Analysis with Truly Nontreated Lines Indicated as Treated Lines

Note. Capped vertical lines represent 95% confidence intervals using the $T(G - L)$ distribution. Solid markers denote the nonplacebo tests.
Table 1  Descriptive Statistics of Workers in Treated and Nontreated Lines

<table>
<thead>
<tr>
<th></th>
<th>Treated Lines</th>
<th>Nontreated Lines</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of workers on same line</td>
<td>245.9</td>
<td>242.1</td>
<td>3.770</td>
</tr>
<tr>
<td></td>
<td>(159.5)</td>
<td>(155.8)</td>
<td></td>
</tr>
<tr>
<td>Share high skilled</td>
<td>0.0121</td>
<td>0.0131</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.114)</td>
<td></td>
</tr>
<tr>
<td>Share female</td>
<td>0.933</td>
<td>0.908</td>
<td>0.024*</td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
<td>(0.289)</td>
<td></td>
</tr>
<tr>
<td>Share from Karnataka</td>
<td>0.858</td>
<td>0.837</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.349)</td>
<td>(0.369)</td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td>28.43</td>
<td>28.40</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(6.808)</td>
<td>(6.484)</td>
<td></td>
</tr>
<tr>
<td>Tenure in plant (in years)</td>
<td>1.117</td>
<td>1.013</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>(2.005)</td>
<td>(1.670)</td>
<td></td>
</tr>
<tr>
<td>Number of Workers</td>
<td>579</td>
<td>1068</td>
<td></td>
</tr>
</tbody>
</table>

Note. Standard deviations in parentheses
* p<0.1, ** p<0.05, *** p<0.01
### Table 2  Effect of Monitoring on Line Productivity (Percent Efficiency)

<table>
<thead>
<tr>
<th></th>
<th>Before RFID Implemented</th>
<th>After RFID Implemented</th>
<th>Difference: After—Before</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Panel 1: Unconditional Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Across All Lines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>65.9% (0.2)</td>
<td></td>
<td>70.0% (0.2)</td>
</tr>
<tr>
<td>Control</td>
<td>59.0% (0.2)</td>
<td></td>
<td>60.6% (0.2)</td>
</tr>
<tr>
<td>Treated—Control</td>
<td>6.8% (0.3)</td>
<td></td>
<td>9.4% (0.3)</td>
</tr>
<tr>
<td><strong>Panel 2: Analysis by Product</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a. Simple Product Lines (Pants)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>64.2% (0.2)</td>
<td></td>
<td>73.0% (0.2)</td>
</tr>
<tr>
<td>Control</td>
<td>59.4% (0.2)</td>
<td></td>
<td>60.1% (0.3)</td>
</tr>
<tr>
<td>Treated—Control</td>
<td>4.8% (0.3)</td>
<td></td>
<td>12.9% (0.4)</td>
</tr>
<tr>
<td>2b. Complex Product Lines (Jackets)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>68.8% (0.4)</td>
<td></td>
<td>63.9% (0.4)</td>
</tr>
<tr>
<td>Control</td>
<td>57.5% (0.4)</td>
<td></td>
<td>62.1% (0.3)</td>
</tr>
<tr>
<td>Treated—Control</td>
<td>11.2% (0.5)</td>
<td></td>
<td>1.8% (0.6)</td>
</tr>
<tr>
<td><strong>Panel 3: Analysis by Operation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a. Simple Operations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>60.0% (1.2)</td>
<td></td>
<td>65.1% (1.2)</td>
</tr>
<tr>
<td>Control</td>
<td>57.1% (1.2)</td>
<td></td>
<td>57.2% (0.7)</td>
</tr>
<tr>
<td>Treated—Control</td>
<td>2.9% (1.8)</td>
<td></td>
<td>7.9% (1.5)</td>
</tr>
<tr>
<td>3b. Complex Operations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>71.6% (1.4)</td>
<td></td>
<td>66.4% (1.3)</td>
</tr>
<tr>
<td>Control</td>
<td>65.4% (1.0)</td>
<td></td>
<td>64.4% (0.7)</td>
</tr>
<tr>
<td>Treated—Control</td>
<td>6.2% (1.7)</td>
<td></td>
<td>2.0% (1.5)</td>
</tr>
</tbody>
</table>

*Note.* Cells represent mean percent efficiency at the line-day level (1, 2a, and 2b) and operation-day level (3a and 3b). Analysis 2a is for lines producing the product with relatively simple operations (pants), 2b is for lines producing the product with relatively complex operations (jackets), 3a is for relatively simple operations on the pant lines, and 3b is for relatively complex operations on the pant lines.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>1.60</td>
<td>1.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.08)</td>
<td>(1.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>6.82**</td>
<td>6.85**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.37)</td>
<td>(2.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Treated</td>
<td>2.57</td>
<td>2.65</td>
<td>2.83</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>(4.46)</td>
<td>(4.30)</td>
<td>(4.40)</td>
<td>(4.21)</td>
</tr>
<tr>
<td>Line FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month-Year FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Day of Week FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>16,436</td>
<td>16,436</td>
<td>16,436</td>
<td>16,436</td>
</tr>
</tbody>
</table>

*Note.* Observations are at the line-date level.

All estimates are from OLS models.

DV: One point is one percent efficiency.

Post=1 after RFID implemented.

Treated=1 for lines that received RFID.

Standard errors clustered by line are in parentheses.

* p<0.1, ** p<0.05, *** p<0.01
Table 4  Difference-in-Differences Regression for Effect of Monitoring Treatment on Efficiency, by Line Complexity

<table>
<thead>
<tr>
<th></th>
<th>Simple Lines (Pants)</th>
<th>Complex Lines (Jackets)</th>
<th>All Lines (Pants &amp; Jackets)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Post</td>
<td>0.73</td>
<td>4.52</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>(2.50)</td>
<td>(2.86)</td>
<td>(2.46)</td>
</tr>
<tr>
<td>Treated</td>
<td>4.80*</td>
<td>11.2***</td>
<td>4.80*</td>
</tr>
<tr>
<td></td>
<td>(2.52)</td>
<td>(0.31)</td>
<td>(2.48)</td>
</tr>
<tr>
<td>Post × Treated</td>
<td>8.06***</td>
<td>-9.42*</td>
<td>-7.79*</td>
</tr>
<tr>
<td></td>
<td>(2.68)</td>
<td>(2.86)</td>
<td>(1.38)</td>
</tr>
<tr>
<td>Complex</td>
<td>-1.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Complex</td>
<td>3.80</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.46)</td>
<td>(2.96)</td>
<td></td>
</tr>
<tr>
<td>Treated × Complex</td>
<td>6.42**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Treated × Complex</td>
<td>-17.5***</td>
<td>-15.7***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.59)</td>
<td>(2.90)</td>
<td></td>
</tr>
</tbody>
</table>

Line FE | No  | Yes | No  | Yes | No  | Yes |
Month-Year FE | No  | Yes | No  | Yes | No  | Yes |
Day of Week FE | No  | Yes | No  | Yes | No  | Yes |
N       | 12,137 | 12,137 | 4,299 | 4,299 | 16,436 | 16,436 |

Note. Observations are at the line-date level. All estimates are from OLS models.
DV: One point is one percent efficiency. Post=1 after RFID implemented.
Treated=1 for lines that received RFID. Jackets=1 for jacket lines.
Standard errors clustered by line are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01
Table 5  Difference in Differences Regression for Effect of Monitoring Treatment on Efficiency, by Operational Complexity in Pant Lines in 2012

<table>
<thead>
<tr>
<th></th>
<th>Simple Operations (Within Pants)</th>
<th>Complex Operations (Within Pants)</th>
<th>All Operations (Within Pants)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Post</td>
<td>0.10</td>
<td>-1.04</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.87)</td>
<td>(1.66)</td>
<td>(0.87)</td>
</tr>
<tr>
<td>Treated</td>
<td>2.93</td>
<td>6.21*</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(3.17)</td>
<td>(2.62)</td>
</tr>
<tr>
<td>Post × Treated</td>
<td>4.97*</td>
<td>5.36*</td>
<td>4.97*</td>
</tr>
<tr>
<td></td>
<td>(2.77)</td>
<td>(3.07)</td>
<td>(2.76)</td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
<td>8.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(7.33)</td>
</tr>
<tr>
<td>Post × Complex</td>
<td>-1.14</td>
<td>-1.12</td>
<td>-1.14</td>
</tr>
<tr>
<td></td>
<td>(1.80)</td>
<td>(1.77)</td>
<td>(1.77)</td>
</tr>
<tr>
<td>Treated × Complex</td>
<td>3.29</td>
<td>1.82</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>(3.98)</td>
<td>(3.84)</td>
<td>(3.98)</td>
</tr>
<tr>
<td>Post × Treated × Complex</td>
<td>-9.13***</td>
<td>-6.36*</td>
<td>-9.13***</td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
<td>(3.58)</td>
<td>(3.11)</td>
</tr>
<tr>
<td>Line FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Month FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Day of Week FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3,745</td>
<td>3,745</td>
<td>1,618</td>
</tr>
</tbody>
</table>

Note. Observations are at the operation-date level. All estimates are from OLS models.

DV: One point is one percent efficiency. Post=1 after RFID implemented.
Treated=1 for lines that received RFID. Complex=1 for SMV greater than 0.75.
Standard errors clustered by operation are in parentheses.

* p<0.1, ** p<0.05, *** p<0.01