Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Discretionary Task Ordering: Queue Management in Radiological Services

http://orcid.org/0000-0002-0261-5397Maria R. Ibanez, Jonathan R. Clark, Robert S. Huckman, http://orcid.org/0000-0002-2674-5831Bradley R. Staats

To cite this article:
http://orcid.org/0000-0002-0261-5397Maria R. Ibanez, Jonathan R. Clark, Robert S. Huckman, http://orcid.org/0000-0002-2674-5831Bradley R. Staats (2017) Discretionary Task Ordering: Queue Management in Radiological Services. Management Science

Published online in Articles in Advance 30 Aug 2017

. https://doi.org/10.1287/mnsc.2017.2810

http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Discretionary Task Ordering: Queue Management in Radiological Services

Maria R. Ibanez,[a] Jonathan R. Clark,[b] Robert S. Huckman,[a] Bradley R. Staats[c]

[a] Harvard Business School, Boston, Massachusetts 02163; [b] University of Texas at San Antonio, San Antonio, Texas 78249; [c] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Contact: mibanez@hbs.edu, http://orcid.org/0000-0002-0261-5397 (MRI); jonathan.clark@utsa.edu (JRC); rhuckman@hbs.edu (RSH); bstaats@unc.edu, http://orcid.org/0000-0002-2674-5831 (BRS)

Received: October 19, 2015
Revised: July 30, 2016; December 5, 2016
Accepted: January 18, 2017
Published Online in Articles in Advance: August 30, 2017

https://doi.org/10.1287/mnsc.2017.2810

Abstract. Work scheduling research typically prescribes task sequences implemented by managers. Yet employees often have discretion to deviate from their prescribed sequence. Using data from 2.4 million radiological diagnoses, we find that doctors prioritize similar tasks (batching) and those tasks they expect to complete faster (shortest expected processing time). Moreover, they exercise more discretion as they accumulate experience. Exploiting random assignment of tasks to doctors' queues, instrumental variable models reveal that these deviations erode productivity. This productivity decline lessens as doctors learn from experience. Prioritizing the shortest tasks is particularly detrimental to productivity. Actively grouping similar tasks also reduces productivity, in stark contrast to productivity gains from exogenous grouping, indicating deviation costs outweigh benefits from repetition. By analyzing task completion times, our work highlights the trade-offs between the time required to exercise discretion and the potential gains from doing so, which has implications for how discretion over scheduling should be delegated.

History: Accepted by Serguei Netessine, operations management.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2017.2810.

Keywords: discretion • scheduling • queue • healthcare • learning • experience • decentralization • delegation • behavioral operations

1. Introduction

The scheduling of work is a key driver of operational performance in many settings, including factories (Berman et al. 1997), trucking (Roberti et al. 2015), healthcare (KC and Terwiesch 2009), and financial services (Staats and Gino 2012). Accordingly, a rich line of research investigates task-scheduling policies, identifying optimal schedules that managers can then implement (Pinedo and Yen 1997, Pinedo 2012). In many settings, however, those who execute the tasks often have discretion over the order in which to perform their assigned duties. Yet little is known about the drivers of workers' decisions to exercise such discretion and how scheduling should be managed when discretion exists. In this paper, we consider the operational drivers and implications of the exercise of "discretion over task ordering," defined as an individual's ability to select which task to complete next from a work queue.

Worker discretion can improve system performance (van Donselaar et al. 2010, Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015) but can sometimes enable workers to "choose the 'wrong' task (operationally)" (Boudreau et al. 2003, p. 186). We consider a worker's decision about which task to execute next among a queue of pending, independent tasks; although the assigned order would suggest choosing the first task in the queue, the worker may choose to exercise discretion by selecting a task from the rest of the queue. Hereafter, "deviation" denotes the exercise of discretion over task ordering by selecting a task that is not the next one in the queue. As technological advances are facilitating the delegation and monitoring of decisions made by front-line workers in manufacturing, services, and knowledge work (Pierce et al. 2015), understanding the operational implications of discretion over task sequence is increasingly important.

We address two research questions. First, what are the drivers of deviations? Second, what are their performance implications? To identify the drivers of deviations, we consider the circumstances under which workers are more likely to exert discretion over the order in which they execute tasks. We posit that the ability of workers to identify an alternative task sequence that they perceive as superior to the assigned sequence will depend on the characteristics of the individual as well as the characteristics of the individual's queue of pending tasks. With respect to the individual, we examine the role of worker experience. As for queue characteristics, we examine whether an individual has an opportunity to pursue a shortest expected processing time (SEPT) policy (i.e., select the task in the queue that is expected to be completed most quickly) or a batching policy (i.e., repeat the prior case type) by deviating.

We investigate these questions using data on doctors reading radiological images of different types (e.g., chest x-rays, head computed tomography (CT) scans) at a company at which images are randomly assigned to the individual queues of qualified doctors. These radiologists deviate from the assigned first-in-first-out (FIFO) scheduling policy 42% of the time. Our findings show that doctors deviate more often when they are more experienced, when there is an opportunity to follow a SEPT policy by deviating, and when there is an opportunity to batch by deviating. When doctors deviate, however, their average read time tends to increase by about 13%. Other performance dimensions, including quality, are mostly unaffected. Overall, our calculations suggest that forgoing deviations would have led to faster reading times that could have saved 2,494 hours per year, which would have increased annual profits by 3%.

We also find that different types of deviations have varied effects on performance. First, the deviation penalty is lower when the worker is more experienced. Second, although SEPT may create the illusion of working faster precisely because it selects shorter cases, it tends to impair speed and is a particularly detrimental type of deviation. Third, consistent with theory, prior empirical work, and the beliefs of the radiologists we interviewed, batching is associated with superior performance when it occurs naturally, yet this is not the case when batching results from a deviation because of the search costs and other time costs associated with actively choosing and selecting the case from the queue. Individuals may seek to group their tasks to achieve the benefits of batching, but this may not be worth it if they need to do so themselves; this provides evidence of the potential harmful performance effects of exercising discretion when an individual may underestimate the costs of deviating in relationship to the potential gain.

Our paper makes several contributions to both theory and practice. Our work is among the first to focus attention on the role of discretion over task sequence in queue management, recognizing that workers may choose their own approach to sequencing or prioritizing work. Whether in call centers, software companies, or doctors' offices, technology increasingly allows managers to choose how much discretion to grant employees with respect to the order in which they complete tasks. This element of system design merits greater theoretical and empirical attention to understand its performance implications, and we provide important evidence related to this goal. Second, we identify conditions under which individuals are more likely to deviate from the assigned queue. Examining the role of experience and the contents of the queue in this decision provides insight into the design of work systems. Third, we make important methodological contributions by identifying a novel approach to discover valid instrumental variables by exploiting exogenous queue contents to evaluate discretion in queuing settings. Finally, we evaluate the performance implications of these choices. Although attention has been given to the performance effects of discretion, the efficiency of discretion—which incorporates the time invested to *exercise* it—has been overlooked. Our analysis suggests that there is a cost of exercising discretion, which managers should take into account when evaluating the effects of delegation. This time cost of reorganizing the queue may make queue improvements inefficient, underscoring the value of having a centralized individual perform queue management rather than dividing it across workers. Deviations, even those that lead to batching and would thus be recommended a priori, may have a higher execution cost than the resulting benefit. Although deviation is unlikely to be detrimental to performance in all situations, our findings illustrate that it can be and that managers must carefully evaluate the full operational implications of allowing discretion.

2. Related Literature

A long line of research on scheduling investigates the optimal allocation of scarce resources (e.g., a machine or a worker) to tasks over time (Pinedo 2012). Problems considered include project scheduling (e.g., Goh and Hall 2013), transportation scheduling (e.g., Zhu et al. 2014), appointment scheduling (e.g., Bassamboo and Randhawa 2016, Truong 2015), and workforce scheduling (e.g., Berman et al. 1997). An influential area of research since the 1950s, the optimization problem can have multiple objectives and typically assumes a central planner. Empirical research has studied the effects of task sequence on performance. Among this work, Schultz et al. (2003) provide experimental evidence of the negative effects of work interruptions, showing that changing machines leads to a performance penalty beyond just the time cost of moving locations. Examining data entry clerks, Staats and Gino (2012) find that repeating the same task is associated with superior shift performance, suggesting that managers should provide variety across days or weeks but minimize task switches within a day.

In this work, the often-implicit assumption is that scheduling is a managerial decision and that workers will execute the schedule chosen by the central planner. In many settings, however, this is not the case; front-line workers have autonomy regarding which task to complete next and, therefore, can deviate from the assigned task schedule. We consider the role of worker discretion with respect to task sequencing. If the exercise of discretion by workers yields better performance, managers should encourage such behavior.

At the same time, if there are costs of exercising discretion, managers should look for ways to lessen these negative outcomes. It is thus important to understand how workers behave when given the freedom to deviate from an assigned task order. Managers should know whether workers deviate frequently and in predictable ways and, if so, whether the choices add value. Our paper addresses these questions.

Although little is known about discretion over task sequence, research has examined discretion with respect to other work dimensions, including capacity allocation (Kim et al. 2015), routing a task to a specialist (Shumsky and Pinker 2003, Saghafian et al. 2014, Freeman et al. 2016), processing time (Schultz et al. 1998, 1999), and balancing the speed–quality trade-off when quality increases with the duration of the interaction (Hopp et al. 2007, Anand et al. 2011, Powell et al. 2012). These studies show that worker discretion has important operational implications (Lu et al. 2014, Tan and Netessine 2014, Berry Jaeker and Tucker 2017) and can help improve system performance (Kim et al. 2015). Research has also examined when individuals make decisions different from those that analytical models assume or recommend. Sometimes these deviations are suboptimal and indicate bias (for reviews, see Bendoly et al. 2006 and Gino and Pisano 2008), such as those identified in inventory management (Schweitzer and Cachon 2000), forecasting (Kremer et al. 2011), or contract structure (Davis et al. 2014). Despite the potential for bias, the management coefficients theory (Bowman 1963) postulates that managers should be allowed to modify decision rules periodically because managers possess valuable information regarding the current environment. For example, van Donselaar et al. (2010) find that store managers at a supermarket chain deviate from the automated inventory order recommendations because of system inadequacy and misaligned incentives, and these deviations add value by diminishing the costs of managing workload and stock-outs. We contribute to and extend this line of work by studying discretion over a different operational variable (scheduling), by introducing task and individual dimensions that may lead to deviation, and by incorporating the time cost of making decisions. By evaluating the efficiency, not just the effectiveness, of discretion, our work highlights the trade-offs between the additional time required to exercise discretion and the potential gains from doing so and enables us to understand better the role of discretion in worker productivity.

3. Discretionary Task Ordering

3.1. Drivers of Deviations from the Assigned Task Order

Many jobs consist of executing a series of sequential, independent, and previously ordered tasks. Examples include doctors seeing patients, mechanics fixing cars, or back-office processors completing claims. In such settings, workers often have visibility into the queue and the ability to deviate from its assigned order, resulting in discretionary task ordering. Although the assigned sequence would suggest choosing the next task in the queue, they may choose a task from the rest of the queue. We refer to the selection of a task other than the next as a "deviation." Workers may choose to exercise discretion over task ordering when they believe that the assigned order is not optimal for performance.[1] We posit that the tendency to deviate will depend on attributes of the worker and the queue.

With respect to the attributes of the worker, we focus on an individual's work experience. First, with experience may come the ability to identify queue inefficiencies and opportunities to improve upon the assigned order. Significant attention has been given to the relationship between experience and process improvement, finding that additional experience typically leads to learning (Lapré and Nembhard 2010, Argote and Miron-Spektor 2011). One reason why experienced individuals show improvement may be that they recognize more opportunities to change their work by altering the order in which they complete tasks. Accordingly, as long as the assigned task sequence is not optimal, individuals should deviate more often as they gain experience. Second, in addition to identifying more improvement opportunities, individuals may be more likely to act upon such opportunities as they acquire more experience. Notably, workers gain confidence through experience (Bandura 1977), and this confidence could encourage them to take action. These arguments lead to our first hypothesis.

Hypothesis 1. *The probability that an individual deviates from the next task in the queue increases with that individual's level of experience.*

We next turn to the attributes of the queue. One task-sequencing strategy that individuals may pursue is a SEPT policy in which the task that is expected to take the least time to perform is completed next. There are operational and behavioral reasons to follow this policy. Operationally, this scheduling discipline minimizes the average number of jobs in the system and the average job wait time. Because of its operational benefits, workers might opt to improve on these dimensions even if these metrics are not used to evaluate performance. Behaviorally, individuals may exhibit a preference for completing easier tasks first even in settings in which their self-interest would be better served by completing tasks in a different order. For example, Amar et al. (2011) find that individuals choose to pay back smaller debts with lower interest rates to accomplish completion instead of paying back the same amount of a larger debt with a higher interest

rate (and thus saving money). In terms of scheduling policies, that means that individuals might first take on what are expected to be the shortest tasks. If workers reorganize the queue according to SEPT, they will be more likely to deviate when the remaining queue (i.e., the queue excluding the first item) contains the task type in the queue with the shortest expected processing time.

Hypothesis 2. *The probability that an individual deviates from the next task in the queue is higher when that task is not of the shortest type in the queue.*

A second category of task-sequencing strategies involves batching—grouping tasks by their types to increase the repetition of similar activities. After completing a task, an individual could recognize that a task further in the queue is very similar to the just-completed task and so choose to complete it next. This batching could bring benefits in terms of decreased setup time, even if the required setup is just cognitive (Staats and Gino 2012). Furthermore, it could also provide processing time benefits, as the relevant knowledge is still in the individual's working memory, allowing the individual to complete the work quickly and avoid interruptions (Bendoly et al. 2014, Froehle and White 2014, Gurvich et al. 2017). Thus, one specific reason to deviate may be the batching of tasks. Batching, however, is not always possible; the opportunity to batch depends on the availability of at least one task of the same type as the predecessor. When the first task in the queue is of the same type as the predecessor, respecting the assigned order would automatically bring the benefits of batching. When the first task in the queue is *not* of the same type as the predecessor, one could deviate toward batching if the rest of the queue contains a repetition of the previous task type. If individuals have an intention to batch, then they will be more likely to deviate when they would not batch by following the first task in the queue but can batch by deviating.

Hypothesis 3. *The probability that an individual deviates from the next task in a queue is higher when that task is different from its predecessor and the remainder of the queue offers an opportunity to repeat the predecessor task type* (*i.e., to batch*).

3.2. Performance Implications

At least since the origin of the scientific management movement (Taylor 1911), the field of operations has sought to understand the drivers of performance. Over time, improvements have been identified in many areas, from task scheduling (Pinedo and Yen 1997) to product variety (Fisher and Ittner 1999) to queuing system design (Gans et al. 2003). Research has increasingly considered human aspects of operations management problems (Boudreau et al. 2003), incorporating such behavioral factors as workload (KC and Terwiesch 2009, Powell et al. 2012, Tan and Netessine 2014, Berry Jaeker and Tucker 2017) and team composition (Huckman et al. 2009, Schultz et al. 2010). Recent work shows that, under certain conditions, expert discretion over operational variables can improve decisions (Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015). In this paper, we study how discretion over task sequence affects task completion time. The completion time for a person carrying out a task is given by the setup time plus the processing time. Under discretionary task ordering, by which individuals select which task to execute next, setup time includes both the search time investigating the queue and choosing a task (task selection time) and the time preparing to execute the new task (standard setup or changeover). Processing time represents the "run time" required to complete the task itself.

We begin by exploring how the exercise of discretion might improve completion time. To the extent that individuals deviate to enhance task sequence, we would expect the exercise of discretion to benefit productivity. Front-line personnel often have information about improvement opportunities that is not available to a central planner (MacDuffie 1997, Tucker 2007, Staats et al. 2011). For example, delivery drivers may adjust their daily route after observing a road under construction during the prior day. Thus, even if we assume that the queue has been optimally organized by the central planner with the knowledge that the planner has, the worker may recognize opportunities to improve upon that plan because of information asymmetry. Moreover, many queues are not optimally organized to begin with, creating more improvement chances. Hence, workers may deviate to apply generally accepted best practices for task scheduling. For example, workers can avoid the cost of switching (e.g., either mental or physical setup costs) by selecting a task that repeats the predecessor's task type (i.e., batching). These improved sequences may thus result in superior speed.

Hypothesis 4A. *Task deviation leads, on average, to faster completion time.*

Although exercising discretion could be beneficial, it might instead prove distracting. First, by searching through the queue to choose the next task to complete, a worker is adding a search cost (task selection time) to the setup time. Second, switching back and forth from searching through the queue to executing tasks could generate cognitive distractions (KC and Staats 2012, Staats and Gino 2012, Froehle and White 2014) and slow the worker more than the gain from the deviation. In addition, the improvement of task sequence may be suboptimal if workers do not look for optimality but rather satisfice—selecting an option meeting their minimum requirements (Simon 1978). Given the potential for conflicting performance effects, we offer the following competing hypothesis.

Hypothesis 4B. *Task deviation leads, on average, to slower completion time.*

Regardless of the average net effect of deviations on speed, different types of deviations may have varied effects. We first consider the heterogeneous effect of deviations across levels of worker experience. Extensive literature in learning-by-doing shows that individuals' performance improves with experience (Huckman and Pisano 2006, Narayanan et al. 2009). One activity that individuals may learn through experience is how to exercise discretion over task sequence. That is, workers may learn about how to deviate more effectively and efficiently as they gain experience, which, in turn, could lead to better deviations in terms of speed performance. For example, they may develop better intuition about which task to work on next or learn how to search through the queue faster to execute a preferred strategy. Hence, for positive net effects of deviations on speed, we would expect the performance benefit to grow with experience, and for negative net effects of deviations on speed, we would expect the performance penalty to be smaller with experience. We thus hypothesize the following.

Hypothesis 5A. *Task deviation leads, on average, to faster completion time when an individual has greater experience.*

Next, we investigate the performance implications of deviations according to the task selected. We categorize task-type deviations based on two dimensions that are consistent with SEPT and batching. Theory does not generate a clear prediction for the direct effect of a SEPT policy on completion time. On one side, research in psychology suggests that completing tasks motivates (Gal and McShane 2012), and hence SEPT could be associated with faster speed. On the other side, individuals might have an expectation regarding the appropriate time to be working on any given task regardless of its actual complexity. If this expectation regarding how long a task should take is affected by the other tasks in their queues, then when selecting the shortest task among a given queue, workers might allow their processing time to expand beyond or relative to the expected processing time for this particular task type (Hasija et al. 2010). Hence, when selecting the shortest task (following SEPT), they might underadjust their expectation regarding the reasonable time to be working on such task, leading to a slower speed after controlling for the actual complexity of the task. Therefore, following a SEPT policy may affect efficiency either positively or negatively. Disentangling these effects is ultimately an empirical question.

What would be the performance consequences of deviations toward SEPT? Although the direct effect of SEPT may be theoretically unclear, deviations toward SEPT are likely performance-decreasing. In terms of types of deviations, there are reasons to believe that deviations toward SEPT may be particularly time-consuming because the search cost of exploring the queue not only includes looking at the mix of tasks but also includes making a mental estimation of the reading time of each task type and a comparison of those expected times across all tasks. For example, while deviations toward batching only require an individual to search through the queue until a particular task type is identified, SEPT involves going through the *entire* queue, determining the expected processing time of each task type and contrasting it to the shortest one identified to that point. In addition, precisely because deviating toward SEPT requires this consideration of expected processing times for all tasks in the queue, it may magnify the salience of those other (longer) tasks in the queue as reference points (Hossain and List 2012), leading individuals to increase their expectation for how long a task should take and to subsequently expand the actual processing time to fill this time (Hasija et al. 2010). Thus, we expect deviations to be worse for performance when they are toward SEPT than when they are not.

Hypothesis 5B. *Task deviation leads, on average, to slower completion time when the selected task is of the task type with the shortest expected processing time in the queue.*

The final dimension for categorizing deviations is whether they are toward task-type repetition (i.e., batching) or not. To the extent that repetition of task type is associated with superior performance, one would expect deviations to be more effective when they result in batching than otherwise. Moreover, controlling for the direct effect of repetition on speed, deviations toward repetition are expected to be relatively more efficient. Compared with other sequencing strategies that require evaluating the whole queue, contrasting different options, or computing certain metrics (e.g., expected processing time), the strategy based on task-type repetition only requires searching the queue to find a specific type, and this search stops as soon as the first task meeting this requirement is found. Thus, we expect deviations to be more beneficial when they are toward task-type repetition (i.e., when the selected task is of the same type as the predecessor) than when they are not.

Hypothesis 5C. *Task deviation leads, on average, to faster completion time when there is a task-type repetition.*

4. Setting, Data, and Models

4.1. Empirical Setting—Outsourced Teleradiology Services

We test our hypotheses using transaction-level data from one of the largest outsourced radiological services (teleradiology) firms in the United States. In this

setting, radiologists seated at computer workstations—at home or at a reading center—sequentially interpret "cases," each of which corresponds to a set of digital images of a particular technology and anatomical area for a patient. Technologies used in our setting include x-rays (electromagnetic waves; 3.68% of our final sample), CT (84.26%), nuclear medicine (1.01%), magnetic resonance imaging (MRI; 0.85%), and ultrasound (10.20%). The anatomical areas include the abdomen (5.58%), body (combination of areas; 36.31%), brain (33.23%), breast (0.01%), cardio (0.2%), chest (12.78%), gastrointestinal/genitourinary (1.24%), head and neck (2.44%), musculoskeletal (1.31%), obstetrics (2.6%), pelvis (0.53%), spine (3.73%), and other (0.04%). The company receives cases from clients—typically hospitals or physician group practices—and assigns the reading of them to individual radiologists on a round-robin basis following a computer-based algorithm. To be eligible to receive a case, a radiologist must be trained in the technology and anatomical area, licensed by the state, and credentialed by the hospital where the radiological image was created. Given these requirements, most radiologists in the company can interpret the majority of cases, are licensed in over 35 states, and are credentialed at one-third of the client hospitals. Finally, an eligible radiologist must be on duty and not too backed up when the study arrives. Conditional on a radiologist meeting the availability and eligibility requirements, case assignment to radiologists is random.

At any point in time, the radiologists see their own queue of pending cases. We observe—and control for—the factors the radiologist observes when deciding which case to select next. In particular, for each case in the queue, the radiologist sees the time the case was assigned to the queue, the technology employed to create the images, the anatomical location of the study, and the number of images. Once a case is assigned to a particular radiologist, it is not reallocated to another radiologist. Radiologists work independently without supervision and only see cases assigned to them. Because new cases are continually added to this dynamic queue while radiologists are on duty, radiologists make a decision, explicitly or implicitly, regarding which case to read next every time they start a case rather than reordering all cases at the beginning of a shift, which could happen in settings where all tasks to be completed are known initially.

Management expected radiologists to follow a FIFO policy but did not enforce it, thereby leaving the radiologists free to deviate. Therefore, in this setting, individuals seeking to improve performance are given a task schedule based on the random arrival times of cases and are supposed to follow this order, but they are allowed to adjust it. Each case represents a well-defined task, and radiologists have the freedom to decide which case to work on next. Because the number of cases in a radiologist's queue is 5.6, on average, the radiologist can reasonably inspect the queue to evaluate alternative sequencing strategies. Although the company did not provide access to the radiologists whose work is captured in our data, we interviewed several radiologists at different institutions with similar processes to understand how they approached task deviation. We found that dedicated, individual queues and the freedom to alter task sequence were common. These radiologists also indicated that they often chose to deviate from their assigned task ordering. They provided different explanations, including a desire to read faster cases first and an intention to repeat the same technology–anatomy combination (batching), in line with Hypotheses 2 and 3. In particular, the radiologists indicated that they thought batching was helpful in the interpretation of images and highlighted the importance of focusing their attention, if possible, on a specific anatomical area at a given point in time. In our final sample described below, half of the deviations are consistent with these two reasons to deviate. First, 48% of deviations are toward a case of the shortest type available, consistent with a SEPT scheduling policy. Second, 15% of deviations are toward a case-type repetition. Deviating toward batching is not always possible because it is conditional on the case types available in the remainder of the queue. When batching is possible, the percentage of deviations consistent with batching is 46%. Of the cases interpreted, 46% correspond to SEPT, and 13% are case-type repetitions.

The radiologists in our setting aim to maximize their overall speed, subject to delivering the correct clinical interpretation. They seek to maximize speed for both business and clinical reasons. On the business side, teleradiology companies compete for business on the promise of fast service. On the clinical side, unlike some service settings, completion speed is a major determinant of quality, as timely access to the reading report is often critical for the patient's referring doctor to deliver proper treatment. This positive relationship between speed and quality in healthcare has been noted in prior work (Pisano et al. 2001). While radiologists seek to maximize speed, they do so subject to the constraint of maintaining acceptable quality. The teleradiology company tracks reading discrepancies, whereby a customer receiving a radiological report may raise an objection, including minor comments. Such discrepancies are rare, affecting only 0.3% of the images in our final sample, so it is reasonable to assume that the clinical quality of the reads is deemed acceptable.

Numerous features of this research site make it an ideal setting to explore our questions and mitigate concerns about gaming behavior or other reasons unrelated to performance that might cause radiologists to

prioritize certain cases. First, in this teleradiology company, there is no preemption. Both the response and reading times are quick, so radiologists do not interrupt the reading of a case once it is in progress. Second, all cases are deemed urgent, so there is no prioritization based on medical emergency. Third, given the time-sensitivity of the service, cases are not left in a doctor's queue before any break longer than 30 minutes or by the end of their shift; therefore, postponing a job does not affect the case mix or the workload, and there is no need to prioritize shorter cases due to time constraints. Fourth, the type (technology–anatomy) corresponding to arriving cases is independent of doctors' speed and the types of cases they have previously received or completed, addressing any remaining concern about prioritizing cases by type to affect case mix. Fifth, there are no financial incentives to prioritize cases. Radiologists are compensated based on hours worked and must complete each and every case in their queue within the shift.[2] Sixth, there are no mandatory order restrictions (such as required predecessor tasks) or external factors (such as collaboration with other individuals; Halsted and Froehle 2008, Gurvich et al. 2017) that could limit the doctor's discretion. Finally, there is no need to reorder cases to put the images for a single patient together because those images are grouped into a single case prior to entering a radiologist's queue.

4.2. Data

Our data cover all 2,766,209 cases processed by the teleradiology company between July 2005 and December 2007. We observe both the order in which the jobs are assigned to radiologists and the order in which they are completed, with differences in the two being due to a radiologist's exercise of discretion. For each case, we observe its characteristics (e.g., technology, anatomy, number of images), the radiologist who interpreted it, the time it was assigned to the radiologist, and the time it was completed. We then reconstruct the set of cases in a radiologist's queue at any point in time.

We impose three restrictions on the initial sample. First, because we seek to study decisions made by radiologists about whether to deviate from the queue, we limit the sample to those cases that were selected from a queue of at least two cases. This restriction eliminates cases that were the only ones in the queue when the radiologist started them, as there would be no potential for a radiologist to deviate in such instances. Second, we drop cases for which the time elapsed since the last case exceeds 30 minutes (the 99.5th percentile) because we assume that it represents a new shift or break. We do not have records of the exact breaks taken by the radiologists, who lack fixed schedules and rules for breaks. Through this restriction, we define a shift break as a period longer than 30 minutes without completing a case and drop the first observation for each shift because the estimated reading time would include initial setups. No case is left in the queue at the end of these shifts, which is consistent with the company practice of zero backlogs between shifts to ensure quality of care. Our results are robust to alternative cutoffs for shift breaks (20, 40, 60, 90, or 480 minutes). Third, we drop the observations corresponding to four radiologists, each of whom had less than 100 cases in the remaining sample. This facilitates convergence of the model estimation and ensures that the fixed effects do not introduce bias into unconditional probit estimates (described in the Econometric Models section). The final sample for our regression analysis includes 2,408,218 cases of 53 unique case types interpreted by 91 radiologists. Table 1 displays summary statistics and correlations.

4.2.1. Dependent Variables

Deviation from the Assigned Queue Order. *DEVIATION* is a binary variable indicating whether the individual deviated from the assigned order when selecting the current case. This variable takes the value of 1 if the job selected was not the first one in the queue and 0 otherwise. In our sample, doctors deviated from the order in which the cases were assigned to them 42% of the time. The radiologists who deviate the most and the least deviated 59% and 24% of the time, respectively.

Completion (Read) Time. We capture performance using the amount of time the radiologist spends reading a case (total time to select and interpret a case). We estimate this reading time (*READTIME*) for a case as the difference between the time when the current case and the prior case for the radiologist are completed. When a case is not available in the queue when the radiologist submits the previous case, we use the time difference between when the case becomes available (i.e., the case is assigned to the radiologist) and when it is completed (i.e., the radiologist submits the report). This calculation could overestimate reading time for the first case in a shift; because of this, we ignore the first case of each shift for each radiologist. If cases are left in the queue between shifts, our method for calculating reading time could overestimate reading times; however, in our setting, no case is available in the queue when the last case of a shift is completed.

The time stamp is at the minute level; that is, for each case, we know the minute in which it is assigned to the radiologist and the minute in which the radiologist completes the reading. When a case exits the system within the same minute it enters or the prior case is submitted, the time difference has a zero value. Because our empirical models use the natural log of read time, and the logarithm of zero is not defined, we add 1 to all values (Allcott and Sweeney 2017), which is equivalent to rounding up the estimated reading time. The average estimated reading time per case is 3.75 minutes.

Table 1. Descriptive Statistics ($n = 2{,}408{,}218$)

Variable	Definition	Mean	Std. dev.	Min	Max
(1) *READTIME* (minutes)	Amount of time (in minutes) spent working on the current case; this is the completion time	3.755	3.812	0	30
(2) *DEVIATION* (indicator)	Whether the worker deviates from the assigned order and selects a case other than the first one in the queue	0.419	0.493	0	1
(3) *QSIZE* (count)	Length of radiologist's queue (i.e., count of pending jobs) when selecting the current case to be read next	5.553	3.791	2	58
(4) *QVARIETY* (index 0-1)	Variety of case types in the queue, measured as 1 minus the Herfindahl index of different case types in the queue	0.556	0.198	0	0.918
(5) *NUM_IMAGES* (count)	Number of images in the current case	1.414	0.598	1	17
(6) *ORDER_IN_SHIFT* (count)	Case number in the shift—the number of cases read by the radiologist since the shift start, including the current case	57.653	51.612	2	459
(7) *RESTART* (indicator)	Whether the queue was empty when the previous case was finished	0.013	0.114	0	1
(8) *EXPERIENCE* (years)	Employee job tenure in years (with decimals), measured from the number of days that the radiologist has been working at the firm when interpreting the current case	1.906	1.234	0	5.501
(9) *SEPT_OPPTY* (indicator)	Whether the first case in the queue is *not* of the SEPT type within the queue	0.635	0.481	0	1
(10) *SEPT* (indicator)	Whether the current case is of the SEPT type within the queue	0.455	0.498	0	1
(11) *REPEAT_OPPTY* (indicator)	Whether the first case in the queue is different from the case just finished by this radiologist, but it is possible to repeat prior type by choosing another case in the queue	0.206	0.404	0	1
(12) *REPEAT* (indicator)	Whether the current case is of the same type as the case just finished by this radiologist (batching)	0.125	0.331	0	1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(2)	−0.046										
(3)	−0.242	0.200									
(4)	−0.186	0.157	0.430								
(5)	0.182	−0.088	−0.031	−0.239							
(6)	−0.177	0.042	0.159	0.046	0.043						
(7)	0.159	0.136	−0.084	−0.056	−0.032	−0.019					
(8)	−0.054	0.068	0.109	0.071	−0.007	−0.053	−0.013				
(9)	−0.025	0.169	0.213	0.398	0.136	0.048	0.001	0.043			
(10)	−0.050	0.048	−0.189	−0.335	−0.392	−0.056	0.069	−0.014	−0.545		
(11)	−0.065	0.104	0.193	0.226	−0.002	0.007	−0.016	0.039	0.308	−0.155	
(12)	−0.092	0.070	0.024	0.031	−0.257	−0.040	0.011	0.013	−0.116	0.307	0.203

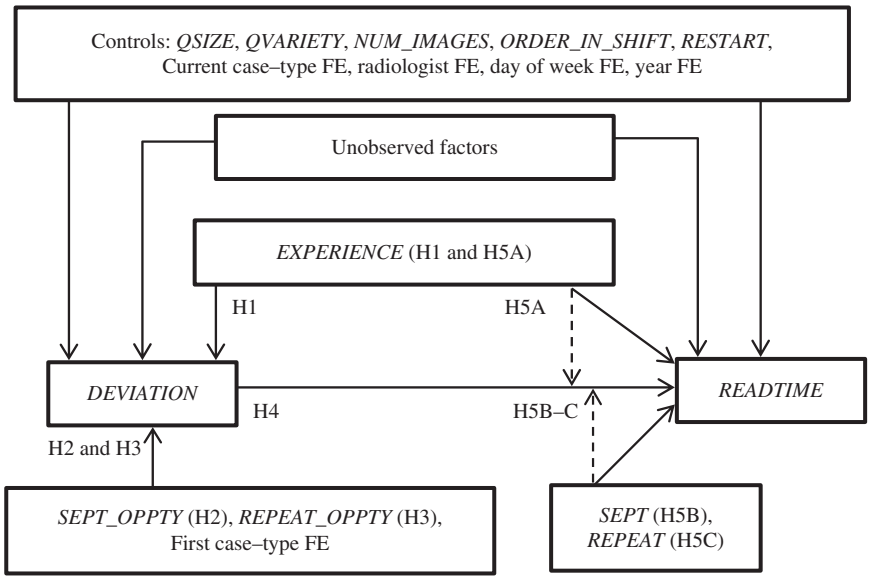
Notes. The unit of analysis is an individual case. Expected processing time is calculated as the average reading time for the given case type (technology and anatomy) for the focal radiologist. A case is a repetition if it is of the same type as the previous case read by the radiologist. We do not consider a repetition of the anatomy categories "body" or "other" as a repetition, since each case in these categories is unique. We compute whether there is a repetition using the full sample before imposing the restrictions described in the Data section.

4.3. Identification Strategy

Our field data allow us to establish external validity, identify effect sizes, overcome observer bias, and study the phenomenon in a rich context over a relatively long time period. One challenge with these data, however, is that the decision to deviate from the assigned order is possibly endogenous. Specifically, there may be unobserved factors that affect both the decision to deviate and performance in terms of the reading time for a case, which is illustrated in Figure 1. Disregarding this endogeneity could lead to bias. We address this challenge by exploiting a set of instrumental variables and estimating an endogenous treatment-regression model (Heckman 1978, Maddala 1983) composed of an equation for the treatment (i.e., the decision to deviate from the first case in the queue) and an equation for performance. The performance equation is identified if and only if at least one exogenous regressor excluded from this equation has a nonzero coefficient in the other (i.e., deviation) equation. This is known as the rank condition. This simultaneous equation system can be interpreted as using instrumental variables for the endogenous regressor (*DEVIATION*). A valid instrument must have two characteristics: first, it must be exogenous—that is, contemporaneously uncorrelated with the error, influencing the outcome (i.e., performance) only through the deviation decision—and second, it must be correlated with the endogenous variable, *DEVIATION*.

We present a novel approach to identify valid instruments to study discretion in queuing settings. Specifically, we propose that the choice to exercise discretion is affected by the composition of the tasks within the queue, and, when exogenous, such composition can

Figure 1. Causal Diagram



Notes. The instrumental variables used to account for the endogeneity of the decision to deviate (*DEVIATION*) on performance (*READTIME*) are *SEPT_OPPTY*, *REPEAT_OPPTY*, and first case–type fixed effects. FE, fixed effects.

generate valid instruments. It is precisely these queue contents that allow the use of discretion, determining the options available to the decision maker. For example, queues with only one item do not allow for discretion regarding which item to select. Queues with only one *type* of item do not allow discretion over which task type to work on. Alternatively, queues that offer a choice of different opportunities (in terms of types of items, sequencing strategies, or other dimensions) provide the opportunity to employ discretion. When queue contents are exogenous, certain queue characteristics may be valid instruments. This approach can be applied to different types of decisions involving queues. For example, the exogenous arrival of a new task in a queue during the processing of a given task could be used as an instrument to study task interruptions and preemption.

We use multiple queue characteristics to instrument for deviation from the next case in the queue. The first instrumental variable is the opportunity to deviate to follow a shortest expected processing time policy (*SEPT_OPPTY*) because of the availability of a shorter case in the remaining queue. To the extent that a radiologist might pursue a SEPT policy, the opportunity to do so by deviating will affect the decision to deviate. *SEPT_OPPTY* is positively correlated with *DEVIATION* (see Table 1). On average, radiologists deviate 48% of the time when there is an opportunity to follow a SEPT policy by deviating, while they only deviate 31% of the time when such opportunity does not exist (see Table 2). Thus, there is a 17-percentage-point difference in the deviation rate between the decisions made when the first case is the shortest in the queue versus when it is not. At the same time, *SEPT_OPPTY* does not directly affect performance.

The second instrumental variable is the opportunity to deviate to repeat the task type (*REPEAT_OPPTY*). *REPEAT_OPPTY* is positively correlated with *DEVIATION* (see Table 1). On average, when there is an opportunity to repeat case type only by deviating, radiologists deviate 52% of the time, while they only deviate 39% of the time when such opportunity does not exist (see Table 2). There is thus a 13-percentage-point increase in deviation associated with the presence of an opportunity to batch by deviating. At the same time, *REPEAT_OPPTY* does not directly affect performance.

Our final set of instrumental variables captures the case type of the first case in the queue. Doctors may be more likely to choose (by deviating or not) certain task types. The more attractive the type of the first case

Table 2. Observed Deviation Rate for Cases Selected Among Different Queues

Queue characteristic used as instrumental variable (IV)	Percentage of deviations if IV = 0 (%)	Percentage of deviations if IV = 1 (%)
Opportunity to follow SEPT (*SEPT_OPPTY*)	30.9	48.3
Opportunity to repeat case type (*REPEAT_OPPTY*)	39.3	52.1
First case–type indicators	41.0–44.8	33.0–78.3

Note. This table describes the percentage of decisions ($n = 2{,}408{,}218$) regarding which case to work on next among queues of different characteristics in which the radiologist deviated from the next case in the queue (*DEVIATION* = 1).

in the queue, the more likely the radiologist would be to select this case next and thus the less likely the radiologist would be to deviate. Therefore, the case type of the first case in the queue is expected to affect the decision to deviate from the queue.[3] At the same time, after controlling for the case type of the selected case, the case type of the first case in the queue does not directly affect performance. This corresponds to the case that was supposed to be read next (according to the assigned order, had the radiologist not deviated) while the speed of the read will only depend on the type of case selected and actually read—a factor for which we control. In conclusion, there is substantial support for the validity of these variables as instruments.

Using this instrumental variables approach, the estimates represent the average performance effect of deviation for the subgroup of task selections affected by these instruments. The identification is driven by a comparison of differences in the reading times among cases with similar observable characteristics (e.g., same type, interpreted by the same radiologist) but for which the doctor deviated or did not deviate to select them because of the different queue characteristics related to (1) opportunities to follow a SEPT policy by deviating, (2) opportunities to repeat case type by deviating, and (3) the default choices as represented by the type of the first case in the queue. These instruments provide variation that is likely to capture many deviations. First, the large number of instruments employed has the advantage of increasing the subset of tasks considered in estimating the effect, which thus represents a larger portion of the overall population of tasks. Second, and more importantly, the type of the first case of the queue affects a broad set of decisions to deviate because any such decision will compare the default option (given by the first task in the queue) with the remaining options (given by the rest of the tasks in the queue). Accordingly, choosing a different task from the first one (i.e., deviating) is equivalent to rejecting the first task. Although these instruments might not apply to all deviation decisions, they have the potential to affect the large majority of decisions.

4.4. Econometric Models

As discussed in the prior section, we examine workers' decisions to deviate from the assigned queue order and the subsequent performance implications by estimating an endogenous treatment regression model (Heckman 1978, Maddala 1983) composed of two equations—a probit model for the treatment (i.e., the decision to deviate) and a log-linear regression model for speed performance:

$$DEVIATION_{ij} = \begin{cases} 1 & \text{if } \mathbf{X}_{ij}\boldsymbol{\beta} + \varepsilon_{ij} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$\ln(READTIME_{ij}) = \delta DEVIATION_{ij} + \mathbf{X}'_{ij}\boldsymbol{\beta}' + u_{ij}. \tag{2}$$

Here, i and j denote radiologists and cases, respectively, and the random disturbance terms ε_{ijt} and u_{ij} are bivariate normal with mean zero and covariance matrix $\left[\begin{smallmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{smallmatrix}\right]$.

The vectors of covariates $\mathbf{X}_{ij}$ and $\mathbf{X}'_{ij}$ include fixed effects for radiologist (to control for time-invariant radiologist characteristics), day of week, calendar year, and case type (defined by the unique combinations of technology and anatomy of the case, to control for heterogeneity in attractiveness and average reading time across case types); the number of years the radiologist has been working at the company (*EXPERIENCE*); and controls for (a) the number of cases in the radiologist's queue when the current case is selected (*QSIZE*), which captures both a radiologist's range of options as well as the workload and is, therefore, expected to affect both the decision to deviate and performance; (b) the case-type variety in the queue (*QVARIETY*), representing alternative options for the individual to choose from; (c) the number of images in the current case (*NUM_IMAGES*), as a larger number of images involves additional reading time and could, therefore, affect the likelihood of deviation; (d) the number of cases read by the radiologist since the beginning of the current shift (*ORDER_IN_SHIFT*), as each additional case contributes to both warm-up and fatigue over the course of the shift; and (e) an indicator for whether the queue was empty when the previous case was finished (*RESTART*), accounting for the warm-up effects after being idle for a short period within a shift. This restarting is infrequent in our sample, occurring in only 1% of cases, and is not included in the deviation model because it perfectly predicts the outcome.

In addition to these common covariates, $\mathbf{X}_{ij}$ in the deviation model includes our instrumental variables: an indicator for whether the first case in the queue does not have the shortest expected processing time within the queue (*SEPT_OPPTY*), an indicator for whether there is an opportunity to repeat case type but only by selecting a case from the queue other than the first one (*REPEAT_OPPTY*), and indicators for the type of the first case in the queue. Finally, in the extended model to test Hypotheses 5A–5C, $\mathbf{X}'_{ij}$ in the performance model also includes (a) the indicator variable *SEPT*, which equals 1 if the case read corresponds to the shortest case type in the queue and 0 otherwise; (b) the indicator variable *REPEAT*, which equals 1 if the prior case was of the same type as the current one and 0 otherwise; and (c) interaction terms of *DEVIATION* with *EXPERIENCE*, *SEPT*, and *REPEAT*.

We note four important points regarding the empirical specification. First, the simultaneous equation system takes into account the fact that the deviation decision is determined within the model rather than

predetermined. Given that *DEVIATION* is potentially endogenous, ordinary least squares should not be applied to estimate the performance equation because the estimators would potentially be not only biased but also inconsistent. The maximum likelihood estimator of the simultaneous equation system presented is consistent (Heckman 1978, Maddala 1983). Second, because we do not have a measure of radiologists' experience prior to joining the company, Equation (2) uses the exponential learning curve model, which prevents bias from our lack of information about prior experience (Lapré and Tsikriktsis 2006). Third, although the maximum likelihood estimator in the presence of fixed effects in discrete choice models shows a finite sample bias when the number (T) of observations per individual is very small (i.e., the incidental parameters problem), this bias declines rapidly as T increases beyond 3 and is negligible for large T (Greene 2004). Given that we have a deep panel with an average of 26,464 cases per radiologist and at least 239 cases per radiologist, the only problem estimating unconditional fixed effects is computational. Finally, although the dependent variable completion time would suggest that a survival model could be used (Lu et al. 2014), we do not have a censoring or truncation problem. Hence, a log-linear model is appropriate.

5. Results and Discussion

5.1. The Determinants of Deviations

To investigate the drivers of deviations from the queue, probit maximum likelihood estimates of Equation (1) are shown in Table 3. Because our dependent variable is binary, we use probit regression, but a linear probability model yields similar inferences. Average marginal effects (AMEs) are provided next to the corresponding coefficients. Standard errors are clustered at the radiologist level. In the baseline model (column (1)), we only include the controls. The estimated coefficient on the length of the queue (*QSIZE*) is positive and significant; it may be that larger queues offer more opportunities to deviate or that they create workload pressures that lead a radiologist to choose to deviate. It is possible that the use of discretion could be part of what leads to the eventual "speed-up" effect observed with larger queues (KC and Terwiesch 2009, Staats and Gino 2012, Delasay et al. 2017). The magnitude of the average marginal effect implies that one more case pending in the queue increases the probability of deviating by 1.84 percentage points (a 4.39% increase when compared with the sample average of 41.9%). Keeping the size of the queue fixed, individuals adhere to the queue order less frequently as the variety of different case types available within the radiologist's queue (*QVARIETY*) goes up; this result is consistent with greater options creating more opportunities for radiologists to exercise their discretion over task scheduling. Specifically, all else constant (including queue length), a one-unit increase in the variety of the queue increases the probability of deviating from the queue by about 15.50 percentage points (a 36.99% increase compared with the sample average of 41.9%). In addition, the coefficient estimate for the number of images (*NUM_IMAGES*) is not statistically significant, while the coefficient estimate for the position within the radiologist's shift (*ORDER_IN_SHIFT*) is negative and statistically significant.

To test the hypothesis that worker characteristics affect adherence to the assigned work order, radiologist *EXPERIENCE* is included in column (2). The results

Table 3. Drivers of Deviations (Dependent Variable = *DEVIATION*)

	(1a) Coefficients	(1b) AME	(2a) Coefficients	(2b) AME	(3a) Coefficients	(3b) AME
QSIZE	0.0524*** (0.0039)	0.0184	0.0521*** (0.0039)	0.0183	0.0512*** (0.0038)	0.0180
QVARIETY	0.4408*** (0.0220)	0.1550	0.4307*** (0.0211)	0.1512	0.3447*** (0.0195)	0.1210
NUM_IMAGES	−0.0103 (0.0100)	−0.0036	−0.0084 (0.0101)	−0.0030	−0.0079 (0.0101)	−0.0028
ORDER_IN_SHIFT	−0.0004*** (0.0001)	−0.0001	−0.0004*** (0.0001)	−0.0001	−0.0004*** (0.0001)	−0.0001
EXPERIENCE			0.2195*** (0.0158)	0.0771	0.2201*** (0.0158)	0.0773
SEPT_OPPTY					0.0685*** (0.0068)	0.0241
REPEAT_OPPTY					0.0436*** (0.0046)	0.0154

Notes. This table reports the results from maximum-likelihood probit estimation. The unit of observation is a radiologist's case. The number of observations is 2,408,218. Robust standard errors (in parentheses) are clustered by radiologist. All specifications include fixed effects for case type (technology and anatomy) of the first case in the queue, case type of the focal case (i.e., the case selected and read), radiologist, day of week, and year.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

show that longer tenure at the company is associated with a higher likelihood of deviation, which supports Hypothesis 1. An additional year of experience increases the probability of deviating by 7.71 percentage points (an 18.40% increase when compared with the sample average of 41.9%). We measure experience by the number of years that the radiologist has worked at the company because it allows us to capture experience prior to our sample period, and it is a common measure in the literature (Tucker et al. 2007). Using a radiologist's case volume as an alternative measure of experience yields similar results.

We next explore the impact of queue characteristics on deviation, looking at two particular deviation strategies—deviation toward the shortest cases and deviation toward batching of case types (column (3)). Including an indicator for whether the first case in the queue is inconsistent with a SEPT policy, thereby creating an opportunity to follow SEPT by deviating (i.e., *SEPT_OPPTY* = 1), we find that the predicted probability of deviating from the assigned order is higher when the first case in the queue is not the shortest case in the queue, supporting Hypothesis 2. The average marginal effect suggests that having an opportunity to follow a SEPT policy by deviating boosts the probability of deviation by 2.41 percentage points, increasing the average predicted probability of deviation from 40.87% to 43.28%. The results also indicate that individuals are more likely to deviate when the first case in the queue is not of the same type as the predecessor but the remainder of the queue offers an opportunity to repeat (i.e., *REPEAT_OPPTY* = 1), which is predicted by Hypothesis 3. The average marginal effect indicates that having an opportunity to batch by deviating from the queue increases the probability of deviation by 1.54 percentage points, increasing the average predicted probability from 42.10% to 43.64%.

5.2. The Impact of Deviations on Performance

To study the impact of deviations on performance, we estimate Equations (1) and (2) jointly via maximum likelihood (column (1) of Table 4). Using a control-function estimator provides equivalent results. Standard errors are clustered by radiologist. The Wald test of independent equations indicates that we can reject the null hypothesis of exogeneity of deviation ($p < 0.0001$). The estimated correlation between the treatment-assignment errors and the outcome errors, ρ, is negative, indicating that unobservables that increase reading time tend to occur with unobservables that reduce deviation occurrence (negative bias). Accounting for the endogeneity of the deviation decision is important in obtaining consistent estimates of the deviation effect on reading times. For each additional case in the queue (*QSIZE*), the average reading time decreases, all else equal, by about 2.9% on average. For a one-unit increase in *QVARIETY*, there is a 20% decrease in completion time. Reading time increases for each additional image (*NUM_IMAGES*) included in the case by about 18.7% and decreases over the course of a given shift (*ORDER_IN_SHIFT*) although by a small amount on a case-by-case level. We find evidence of learning-by-doing; on average, an additional year of *EXPERIENCE* decreases reading time per case by 5.4%, holding all else constant. On average, reading time per case more than doubles after a temporary period of zero workload because of restarting effects (*RESTART*). Turning to our main independent variable, we find that, on average, *DEVIATION* is associated with slower reading times. Compared with cases that are first in queue, cases that are deviations take 13.3% longer on average. This supports Hypothesis 4B rather than Hypothesis 4A in our setting. These results provide evidence of the cost of exercising discretion and call for managerial and academic attention.

Although they tend to worsen performance, deviations from the queue are frequent. Why do individuals take actions that ultimately work against their own interests? To understand this question (and test Hypotheses 5A–5C), we estimate a model that distinguishes the effects of different types of deviations. Column (2) of Table 4 shows the results from maximum likelihood estimation of the simultaneous equation system. The predicted mean reading times per case (in minutes) derived from this model are shown in Table 5.

Worker Experience. We first discuss the heterogeneous effects of deviations by worker experience (see Table 5, panel A). We find that tenure ameliorates the negative effect of deviation on performance, consistent with Hypothesis 5A and suggesting that radiologists may be more efficient at deviating or choose better types of deviations as they become more experienced. At each integer level of years of experience, deviations have a higher predicted mean reading time than cases in which the radiologist follows the assigned queue order, suggesting that learning about how to deviate does not overcome the net cost of deviating in our sample. The predicted mean reading time for deviations after three years at the company is equivalent to the prediction for adherence to the assigned order by newcomers in their first year ($\chi^2(1) = 0.60, p = 0.4405$), indicating that the deviation penalty is large enough to suppress the learning from two years of experience.

The overall effect of experience on the performance impact of deviations depends on how experience affects both the individual impact of each deviation and the frequency with which they occur. To evaluate this, we consider how productivity changes for a radiologist over the course of a year. We combine the results from the deviation and the performance models. A one-year increase in experience from the average (two

Table 4. Performance Implications of Deviations and Task Sequence (Dependent Variable = ln(*READTIME*))

	(1) ML	(2) ML	(3) OLS	(4) OLS
QSIZE	−0.0290***	−0.0280***	−0.0267***	−0.0265***
	(0.0023)	(0.0022)	(0.0020)	(0.0020)
QVARIETY	−0.2230***	−0.1894***	−0.1997***	−0.1643***
	(0.0146)	(0.0141)	(0.0139)	(0.0137)
NUM_IMAGES	0.1714***	0.1711***	0.1711***	0.1710***
	(0.0061)	(0.0062)	(0.0063)	(0.0063)
ORDER_IN_SHIFT	−0.0003***	−0.0004***	−0.0004***	−0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
EXPERIENCE	−0.0560***	−0.0487***	−0.0459***	−0.0461***
	(0.0128)	(0.0124)	(0.0125)	(0.0125)
RESTART	0.8412***	0.8336***	0.8688***	0.8682***
	(0.0195)	(0.0195)	(0.0179)	(0.0179)
DEVIATION	0.1250***	0.0875***		
	(0.0158)	(0.0146)		
EXPERIENCE × *DEVIATION*		−0.0090***		
		(0.0035)		
SEPT		0.0161***		0.0344***
		(0.0051)		(0.0049)
SEPT × *DEVIATION*		0.0373***		
		(0.0047)		
REPEAT		−0.0099**	−0.0170***	−0.0171***
		(0.0047)	(0.0048)	(0.0048)
REPEAT × *DEVIATION*		−0.0170***		
		(0.0048)		
ρ	−0.0850***	−0.0441***		
ρ standard error	(0.0129)	(0.0105)		
Test $\rho = 0$ (p-value)	0.0000	0.0000		

Notes. Columns (1) and (2) report the results from joint maximum-likelihood (ML) estimation of the performance model (probit deviation model with full specification, such as in column (3) of Table 3, not shown). Columns (3) and (4) report the results from ordinary least squares estimation of the performance model. The unit of observation is a radiologist's case. The number of observations is 2,408,218. Robust standard errors (in parentheses) are clustered by radiologist. All models include fixed effects for case type (technology and anatomy) of the focal case (i.e., the case selected and read), radiologist, day of week, and year. The probit models also include fixed effects for the case type of the first case in the queue.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

years) increases the frequency of deviation by 7.71 percentage points, from 41.9% to 49.6%. At the same time, it reduces the penalty associated with each deviation, from a 9% increase in reading time (compared with cases that are first in queue) to an 8% increase. Combined, these results suggest that a one-year increase of experience from the average results in a 5% decrease in reading times, on average. Given that, on average, a radiologist interprets approximately 11,000 cases per year, the time saved during a year by an additional year of experience corresponds to 33 hours. If the deviation tendency had not increased, the one additional year of experience would have delivered a productivity boost of 37 hours (i.e., 4 extra hours). Thus, experience still leads to better performance, but the improvement is lower than what would have happened if the deviation tendency had not increased.

To understand how the contents of the queue affect the performance impact of exercising discretion, we next consider two deviation strategies that depend on queue contents: (1) choosing the shortest case in the queue and (2) repeating case type.

SEPT. According to Table 5, panel B, following SEPT increases the predicted mean reading time by 2% and 5% for nondeviations and deviations, respectively ($(3.83 - 3.77)/3.77 = 0.02$ and $(4.26 - 4.04)/4.04 = 0.05$). The deviation penalty corresponds to 11% and 7% of the reading time when following SEPT and otherwise, respectively, based on the second and first rows of the table ($(4.26 - 3.83)/3.83 = 0.11$ and $(4.04 - 3.77)/3.77 = 0.07$). Therefore, our results suggest that following a SEPT policy hurts performance, in general, and increases the performance penalty of deviations, supporting Hypothesis 5B.

Table 5. Predicted Mean Reading Time (in Minutes) per Case

	No deviation (*DEVIATION* = 0)	Deviation (*DEVIATION* = 1)	Difference
Panel A: By number of years of experience			
One year of *EXPERIENCE*	3.97	4.35	0.39 (10%)
Two years of *EXPERIENCE*	3.78	4.11	0.33 (9%)
Three years of *EXPERIENCE*	3.60	3.88	0.28 (8%)
Four years of *EXPERIENCE*	3.43	3.66	0.23 (7%)
Five years of *EXPERIENCE*	3.27	3.46	0.19 (6%)
Panel B: By whether following a SEPT policy			
Not the shortest case (*SEPT* = 0)	3.77	4.04	0.27 (7%)
Shortest case (*SEPT* = 1)	3.83	4.26	0.43 (11%)
Panel C: By whether repeating the case type of the predecessor			
No repetition (*REPEAT* = 0)	3.80	4.15	0.35 (9%)
Repetition (*REPEAT* = 1)	3.77	4.04	0.27 (7%)
Panel D: By whether following a SEPT policy and/or repeating case type			
Not the shortest case, no repetition	3.78	4.05	0.28 (7%)
Not the shortest case, repetition	3.74	3.95	0.21 (6%)
Shortest case, no repetition	3.84	4.28	0.44 (11%)
Shortest case, repetition	3.80	4.16	0.36 (10%)

Notes. These tables report the predicted mean reading time (in minutes) per case based on the performance model in column (2) of Table 4. The difference within row represents deviation versus nondeviation. The difference within column shows learning (panel A), the effect of SEPT (panel B), the effect of task-type repetition (panel C), and the effects of alternative task sequences in terms of *SEPT* and *REPEAT* (panel D). In panel C, the first column compares repetitions versus nonrepetitions when there is no deviation, and the second column compares repetition versus nonrepetition when there is deviation. In panel D, the first column compares task sequences (in terms of *SEPT* and *REPEAT*) when there is no deviation, and the second column compares task sequences (in terms of *SEPT* and *REPEAT*) when there is deviation.

Repetition of Case Type. A common reason cited by radiologists for deviation is the desire to batch cases of the same type. As discussed above, research documents that task repetition is associated with improved performance; this is the argument behind why an individual would choose to deviate toward batching (Hypothesis 3). To analyze the impact of batching on performance without accounting for the deviation choice (and hence, without a deviation model), we run an ordinary least squares model of completion time. We find that, all else constant, average reading times per case tend to be 1.7% lower for those cases that are repetitions of the prior case type than for those cases that are not repetitions (columns (3) and (4) of Table 4). This confirms that the general finding of batching being associated with improved performance holds in this setting. Given these results, as well as the aforementioned received wisdom of operations management, knowledgeable individuals might conclude that they should use their discretion to increase the number of repetitions—and reduce the number of case type "setups"—by reordering tasks. The question thus becomes whether these deviations toward batching are, in fact, beneficial for performance.

To answer this question, we return to our simultaneous-equations system and compare the predicted mean reading times depending on whether the case is a repeat of the prior case type and whether it represents a deviation from the assigned queue order (see Table 5, panel C). When adhering to the assigned sequence, the predicted mean reading time per case is 1% lower when there is a natural (i.e., without deviation) repetition of case type than when there is neither repetition nor deviation (left column, $(3.77 - 3.8)/3.8 = -0.01$, $\chi^2(1) = 4.49$, $p = 0.0340$). Conditional on deviating, the predicted mean read time is 3% lower when there is a repetition than otherwise (right column, $(4.04 - 4.15)/4.15 = -0.03$, $\chi^2(1) = 19.00$, $p < 0.0001$). Thus, the results corroborate that batching is generally associated with superior performance. Batching tends to hurt performance, however, when it is the result of queue reordering; the predicted mean reading time is 6% higher when the radiologist deviates from the queue to take a case that creates a repetition (bottom right cell) compared with cases in which a radiologist neither deviates nor batches (top left cell, $(4.04 - 3.8)/3.8 = 0.06$, $\chi^2(1) = 18.42$, $p < 0.0001$). Although detrimental, this deviation penalty is smaller than the 9% increase in predicted mean reading time when deviating from the queue by taking a case that is *not* a repetition (top row, $(4.15 - 3.8)/3.8 = 0.09$, $\chi^2(1) = 43.06$, $p < 0.0001$). Overall, these results support Hypothesis 5C and suggest that deviations toward batching are less detrimental than other deviations but still have a negative net effect on performance compared

with adhering to the original order. Hence, contrary to expectations that commonly overlook the costs of exercising discretion, with respect to completion time, we find that radiologists should forgo deviations that are aimed at taking advantage of batching. Radiologists would complete their reads faster, on average, if they did not deviate to take advantage of batching, as the benefit of repetition does not compensate for the cost associated with reordering the queue in this setting.

5.3. Evaluating Alternative Policies

We evaluate the overall impact of deviations on productivity using the predicted reading times from Table 5, panel D to compare the status quo (current deviation policy) with three benchmarks involving no deviations. Each benchmark considers a centralized ordering policy with a different assigned task sequence.

Original Sequence. We estimate the reading time for each case based on the values that *REPEAT* and *SEPT* would have had if the radiologists would have followed the assigned queue order, in which *SEPT* was computed over the average queue (five cases), which includes the current case and the next four cases. The resulting average reading time is 3.79% below that corresponding to the status quo. The improvement in speed would have saved 2,494 hours per year for the company. These time savings translate into 39,434 cases per year of additional reading, or an estimated $451,385 salary savings per year for the company.[4] Translated to the bottom line, these savings would have increased annual profits by 3.1%.

REPEAT and SEPT Sequence. The second benchmark is a policy that batches tasks of the same type together and sequences the resulting batches within the shift in order of increasing expected processing time. Although SEPT is associated with lower performance in our setting, radiologists show a tendency toward this policy. This policy might gain doctors' acceptance because it uses their revealed preferences, hence reducing their desire to deviate from the queue. Compared with the status quo, this policy implies a 2.97% decrease in reading times. Over the course of a year, this represents 1,957 hours saved, 30,691 additional cases read, $354,278 of labor cost savings, and a 2.4% increase in annual profits

REPEAT Without SEPT Sequence. Although a policy that sequences similar tasks together and longer tasks first could be harder to implement, as radiologists exhibit a tendency toward prioritizing those tasks they expect to complete faster, SEPT is generally associated with lower performance in our setting. Hence, on the basis of our empirical results, this policy is expected to be the best. Compared with the status quo, this policy implies a 4.27% decrease in reading times. Over the course of a year, this accounts for 2,809 hours saved, 44,640 additional cases, and an estimated $508,426 of labor cost savings. Translated to the bottom line, these savings would have increased annual profits by 3.5%.

5.4. Other Dimensions of Performance

Our analysis focuses on short-term speed performance—the primary concern of managers in this company. An important question is how deviations affect other dimensions of performance—specifically, longer-term speed, employee turnover, and quality. We find that other performance dimensions are mostly unaffected.

First, we examine the effects of deviations on longer-term speed performance. Using extensions of our full simultaneous-equations model, we find that past deviations (measured as either deviations as a proportion of total cases prior to the current case or deviations as a proportion of total cases prior to the current shift) do not affect current speed. In addition, despite the detrimental instantaneous effect of SEPT on the current reading, deviating toward SEPT could have an effect on subsequent tasks by "alleviating" the work of the radiologist. To explore such a delayed effect of SEPT on subsequent tasks, we look at the effect of past SEPT—measured as lagged *SEPT* (i.e., an indicator for whether the prior case interpreted by this radiologist corresponded to SEPT), two lags, three lags, or the proportion of cases that were consistent with SEPT since the beginning of the shift—on current reading time and find that it is associated with slower speed. Thus, SEPT (and hence deviating toward SEPT) does not help future speed.

Second, we examine employee turnover. Departure is infrequent in our sample, with only 6% of radiologists not interpreting cases by the end of the sample. On the basis of logistic and survival analysis, the radiologists' deviations do not predict whether they depart. Hence, it is reasonable to conclude that turnover remains largely unaffected by task sequencing choice.

Third, we look at the effect of deviations on quality, measured by whether there was a discrepancy found for the case. We find that quality is not affected by deviations. Furthermore, there is no evidence of the two specific task sequence strategies—SEPT and REPEAT—affecting quality.

5.5. Why Deviate When It Hurts

Given these findings, why would individuals deviate from their assigned queue order? In the case of following a SEPT policy, there are likely two main reasons. One is that individuals might pursue an alternative performance metric, perhaps seeking to reduce client waiting time rather than reading time. In our setting, this was not a strategy the company expected

(nor wanted) the radiologists to follow, as the waiting time could be kept under control by adjusting the pool of radiologists on duty. An alternative explanation is that individuals misperceived the implications of following a SEPT approach. We note the negative correlation between *DEVIATION* and *READTIME*; it is not until we control for case–type fixed effects that the relationship becomes positive. Hence, precisely because doctors are switching toward shorter cases when they choose the shortest cases within the queue, a SEPT policy may create the mental illusion of working faster. Conditional on a particular case type being selected, the reading time tends to be higher when it has the shortest expected processing time within the queue, but this outcome may be difficult for the individual to anticipate or observe.

In the case of batching, our analysis illustrates a phenomenon that could disentangle the paradox surrounding the detrimental exercise of discretion for other types of deviations as well. We argue that one plausible explanation is that individuals have the illusion of improving performance by exercising discretion because they underestimate, or fail to consider entirely, the time required to do so (e.g., the time required to look through the queue to determine the exact case to complete next and the related cognitive distraction). This task selection time is an opportunity cost. As such, one might expect these costs to be overlooked, as evaluating opportunity costs requires decision makers to account for unrealized, implicit options (Frederick et al. 2009). Seeking to increase their speed, individuals may exercise discretion in a manner that they believe will improve their workflow but that ends up reducing their speed. These results suggest a possible behavioral challenge, as individuals actively pursue strategies that would in fact have been beneficial for performance were it not for the unrecognized costs of pursuing those strategies (e.g., task selection costs). Our findings provide evidence of the vulnerabilities of self-management and the potential value of using centralized management in settings in which the costs of exercising discretion are particularly high relative to the benefits.

The proposition that individuals may underestimate the costs of exercising discretion—and hence believe that deviations help their productivity even in situations when they are detrimental—may explain why deviations rise with experience. As workers deviate more over time, they might erroneously attribute their performance improvements from learning-by-doing to their exercised discretion. Because they only observe their performance improving over time, they may not realize that their deviation behavior is actually hurting them. That is, learning-by-doing may mask any deteriorating effect of deviations, so individuals may increasingly rely on their discretion over time, even in situations in which exercising that discretion undermines overall improvement.

5.6. Managerial Implications

Our paper contributes not only to the theory of operations management but also to its practice. First, we show that discretion has costs that need to be balanced against its potential benefits. Our finding that deviations are, on average, related to worse performance serves as a warning to managers and workers concerning the costs of exercising certain types of discretion. As noted by a radiologist we interviewed, "I always thought that by reading the easiest cases on the queue first, one might end up reading faster, but it seems that the opposite is true. To lend support to your findings, I have noticed that I read more cases on the days I don't go through my list and choose the order in which to read cases. I do think that incorporating information on how individuals order their task sequences could help speed up the process of reading studies and avoid duplicating time spent on organizing workflow."

Managers should pay attention to the effects of deviations on productivity in their settings. Although an initial task sequence assignment might not be optimal, allowing front-line workers to take an active role in scheduling might not be advisable in settings in which the time required to exercise discretion exceeds the benefits of doing so. In any setting, reorganizing the queue takes time, so managers should look for ways to reduce this time while maintaining the benefits from better ordering of queues. Productive nudges from managers could include recommendations on the task to complete next. Reducing workers' desire or need for task reordering, through education about the costs of deviation, centralized queue management, or more responsive software for task ordering, can be a way for managers to improve productivity. In our setting, adherence to the assigned sequence of tasks could result in a meaningful increase in firm profits of roughly 3%.

Our findings regarding the ability of experience and certain types of deviations to offset—although only partially—the detrimental average effect of deviations on performance suggest that organizations need to take these variables into account in structuring work. In settings such as ours, in which deviations tend to reduce performance, the benefits from experience of senior employees are reduced by their tendency to deviate from the queue more often. In such contexts, managers have the opportunity to improve performance by creating awareness, finding ways to persuade experienced employees to adhere to the queue sequence or be even more thoughtful about when they deviate. More generally, in any setting, managers can collaborate with workers to improve scheduling strategies. For example, workers can identify when they deviate from

the assigned order and why they believe it improves performance. With this knowledge, an individual or organization could test these ideas. Identifying productive deviations could help the individual's productivity and perhaps result in beneficial changes to the organization's recommended task schedule. Finally, managers should consider the time required to exercise discretion when using analytics to inform workplace practices, which is illustrated by the fact that the time required to reorder a queue offsets the beneficial effects of batching in our setting.

6. Conclusion

Although prior literature studies task ordering from the perspective of a central scheduler, those who execute the tasks often have discretion over the actual order in which they are performed. In practice, either by constraint or by choice, the delegation of task-scheduling decisions is common and results in individuals self-scheduling their work. As a result of limited research on discretionary task ordering, little is known about how managers should manage scheduling when such discretion exists. Understanding when and how individuals exercise this discretion informs decisions about system design, whether (or to what extent) to grant discretion, how to nudge behavior toward particular uses of discretion, and how to adjust policies to incorporate responses expected from front-line employees.

We consider this underexplored territory by analyzing the drivers and consequences of exercising discretion over task sequence in a setting in which deviations from the queue are observable. Examining a proprietary data set from a teleradiology company in which radiologists are assigned a queue of cases to interpret but are not restricted to follow that order, we find that individuals are more likely to exercise discretion via deviations from the queue when their experience is greater. This finding is consistent with the view that experience leads to superior ability to identify high-leverage opportunities for deviation and/or higher self-confidence to deviate. Our results highlight both the potential power and the limitations of learning-by-doing. On one hand, an individual can learn over time how to exercise discretion over task sequence more effectively. On the other, this learning may not overcome the costs of exercising discretion, and experienced individuals may also fail to assess those costs appropriately. Thus, deviations may remain unnoticeably detrimental even for experienced individuals. We also show that individuals in our setting have a higher probability of using discretion when doing so creates an opportunity to follow either of two particular strategies—SEPT or batching.

The exercise of discretion via deviations from the queue has a net negative effect in our empirical setting—at least in the short to intermediate term. Doctors often choose the "wrong" tasks (e.g., SEPT, found to be detrimental, despite conflicting theoretical predictions), and even when they choose certain "right" tasks (e.g., repetitions), the resulting benefits are smaller than the time cost of deviating. Deviations when the individual is more experienced or to repeat task type are less detrimental but are still related to worse performance compared with the case of no deviation. That is, deviations harm performance even when they are pursued to take advantage of scheduling strategies that are assumed to be—and in the case of batching, actually are—beneficial to performance. In such cases, the benefits of discretion via deviation may not compensate for the costs of exercising it.

6.1. Contributions

Our study offers five main contributions. First, we analyze the implications of discretionary queue management. Scheduling tasks is a critical determinant of employee and organizational productivity (Pinedo 2012). Most studies on task scheduling examine contexts in which scheduling is solely a managerial decision. This is often not the case in practice, however, as front-line workers frequently have discretion in scheduling. Future analytical work should incorporate the endogeneity of task sequences. Accounting for deviations may more accurately reflect many situations and may lead to unexpected recommendations on how to structure work when workers may choose not to implement prescribed schedules.

Second, we provide evidence of costs of *exercising* discretion that may, in settings such as ours, outweigh the associated benefits. Individuals consistently deviate from the order of tasks within their queues despite the fact that this reordering results in cases taking 13% longer on average. Our investigation of batching suggests an important reason why this behavior occurs. When cases are naturally batched in the queue, completion times are faster. When individuals deviate in a manner that is consistent with batching, however, completion times are slower than in situations in which they do not deviate. Thus, it is possible that individuals believe deviations will improve productivity even though they do not. This creates an opportunity to examine ways to encourage individuals to assess the costs of discretion, find ways to reduce such costs, and deviate less when the costs outweigh the benefits. This may be by shifting to centralized ordering of tasks, avoiding duplication of decisions, or encouraging fewer deviations through nudges and information.

Third, we identify conditions under which an individual is more likely to deviate—namely, when that individual has greater experience and when the queue offers shorter cases, more batching opportunities, more tasks, or a greater variety of tasks. Each of

these results offers important theoretical grounding for system design. Future research should explore alternative reasons to deviate based on principles of operations management and identify additional strategies—advantageous or deleterious—that individuals follow in task selection.

Fourth, we provide evidence of the difficulties of achieving optimal behavior in practice. Analytical models have considered workers using their discretion to alter the number of tasks performed and their processing speeds, assuming they do so to maximize a given utility function. Our findings, however, call into question whether workers are able to make these calculations and thereby contribute to recent empirical work that illustrates suboptimal behavior.[5]

Finally, from a methodological perspective, we present a novel strategy to identify instrumental variables to measure the effects of discretion in queuing settings. We propose that the choice of exercising discretion is affected by the composition of the tasks within the queue, and as long as it can be considered exogenous, such composition can provide valid instruments. Exploiting random assignment of tasks to individual queues, together with variation in queue characteristics, we construct instrumental variables based on the expected duration of the pending tasks, the similarity of tasks with the one just finished, and the type of the task in the first position within the queue. More broadly, our approach can be applied to identify different instruments to estimate the impact of operational decisions related to queues.

6.2. Conclusion

Despite the prevalence of discretion in practice, little is known about how workers use discretion over task sequence and how to manage scheduling when discretion exists. Seeking to fill this gap, we consider the implications of discretionary queue management. We find that reordering queues is common in our setting but tends to have negative implications for performance. We identify conditions that encourage—and document the performance effects of—deviations. Overall, our results highlight the need for both managers and academics to pay careful attention to the use of worker discretion in queue management.

Acknowledgments

The authors gratefully acknowledge helpful suggestions and comments from the editors and reviewers, Xiang Ao, Susie Huang, Scott Lee, Michael Luca, Andrew Marder, Ananth Raman, Ariel Stern, and Michael Toffel. The authors also thank management at the field site for their commitment to this work.

Endnotes

[1] Workers may also choose to reorder tasks based on personal incentives. In this paper, we focus on operational drivers and present an empirical setting without personal incentives in conflict with performance.

[2] In this setting, there is no incentive to prioritize cases based on clients, as the client base is extensive and the company does not systematically offer preferential treatment. Consistent with this, and alleviating concerns regarding prioritization of cases based on preferential treatment of certain hospitals, the results are robust to the inclusion of hospital fixed effects.

[3] The basic assumption behind this set of instruments is that they lead to different propensities to deviate. To confirm this, we regress *DEVIATION* against first case–type fixed effects and a set of controls (see column (3) in Table 3). We strongly reject the hypothesis that the fixed effects for the different types of the first case are the same ($p < 0.0001$).

[4] We use the estimated median hourly wage for a radiologist of $181. (*Source*: Physician—Radiology Hourly Wages, http://www1.salary.com/radiologist-hourly-wages.html, accessed April 27, 2016.) The annual profit comparison uses confidential company data.

[5] We thank an anonymous reviewer for this suggestion.

References

Allcott H, Sweeney RL (2017) The role of sales agents in information disclosure: Evidence from a field experiment. *Management Sci.* 63(1):21–39.

Amar M, Ariely D, Ayal S, Cryder CE, Rick SI (2011) Winning the battle but losing the war: The psychology of debt management. *J. Marketing Res.* 48(SPL):S38–S50.

Anand KS, Paç MF, Veeraraghavan S (2011) Quality–speed conundrum: Trade-offs in customer-intensive services. *Management Sci.* 57(1):40–56.

Argote L, Miron-Spektor E (2011) Organizational learning: From experience to knowledge. *Organ. Sci.* 22(5):1123–1137.

Bandura A (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psych. Rev.* 84(2):191–215.

Bassamboo A, Randhawa RS (2016) Scheduling homogeneous impatient customers. *Management Sci.* 62(7):2129–2147.

Bendoly E, Donohue K, Schultz KL (2006) Behavior in operations management: Assessing recent findings and revisiting old assumptions. *J. Oper. Management* 24(6):737–752.

Bendoly E, Swink M, Simpson WP (2014) Prioritizing and monitoring concurrent project work: Effects on switching behavior. *Production Oper. Management* 23(5):847–860.

Berman O, Larson RC, Pinker E (1997) Scheduling workforce and workflow in a high volume factory. *Management Sci.* 43(2): 158–172.

Berry Jaeker JA, Tucker AL (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient quality. *Management Sci.* 63(4):1042–1062.

Boudreau J, Hopp W, McClain JO, Thomas LJ (2003) On the interface between operations and human resources management. *Manufacturing Service Oper. Management* 5(3):179–202.

Bowman EH (1963) Consistency and optimality in managerial decision making. *Management Sci.* 9(2):310–321.

Campbell D, Frei FX (2011) Market heterogeneity and local capacity decisions in services. *Manufacturing Service Oper. Management* 13(1):2–19.

Davis AM, Katok E, Santamaría N (2014) Push, pull, or both? A behavioral study of how the allocation of inventory risk affects channel efficiency. *Management Sci.* 60(11):2666–2683.

Delasay M, Ingolfsson A, Kolfal B, Schultz K (2017) Load effect on service times. Working paper, School of Business, University of Alberta, Edmonton, Canada.

Fisher ML, Ittner CD (1999) The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Sci.* 45(6):771–786.

Frederick S, Novemsky N, Wang J, Dhar R, Nowlis S (2009) Opportunity cost neglect. *J. Consumer Res.* 36:553–561.

Freeman M, Savva N, Scholtes S (2016) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Sci.*, ePub ahead of print September 2, https://doi.org/10.1287/mnsc.2016.2512.

Froehle CM, White DL (2014) Interruption and forgetting in knowledge-intensive service environments. *Production Oper. Management* 23(4):704–722.

Gal D, McShane BB (2012) Can small victories help win the war? Evidence from consumer debt management. *J. Marketing Res.* 49(4):487–501.

Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* 5(2):79–141.

Gino F, Pisano GP (2008) Toward a theory of behavioral operations. *Manufacturing Service Oper. Management* 10(4):676–691.

Goh J, Hall NG (2013) Total cost control in project management via satisficing. *Management Sci.* 59(6):1354–1372.

Greene W (2004) The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics J.* 7(1):98–119.

Gurvich I, O'Leary K, Van Mieghem JA, Wang L (2017) Collaboration, interruptions and changeover times: Model and empirical study of hospitalist processing times. (August 18), https://ssrn.com/abstract=2616926.

Halsted MJ, Froehle CM (2008) Design, implementation, and assessment of a radiology workflow management system. *Amer. J. Roentgenology* 191(2):321–327.

Hasija S, Pinker E, Shumsky RA (2010) Work expands to fill the time available: Capacity estimation and staffing under Parkinson's Law. *Manufacturing Service Oper. Management* 12(1):1–18.

Heckman JJ (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4):931–959.

Hopp WJ, Iravani SMR, Yuen GY (2007) Operations systems with discretionary task completion. *Management Sci.* 53(1):61–77.

Hossain T, List JA (2012) The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Sci.* 58(12):2151–2167.

Huckman RS, Pisano GP (2006) The firm specificity of individual performance: Evidence from cardiac surgery. *Management Sci.* 52(4):473–488.

Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Sci.* 55(1):85–100.

KC D, Staats BR (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing Service Oper. Management* 14(4):618–633.

KC D, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* 55(9):1486–1498.

Kim S-H, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Sci.* 61(1):19–38.

Kremer M, Moritz B, Siemsen E (2011) Demand forecasting behavior: System neglect and change detection. *Management Sci.* 57(10):1827–1843.

Lapré MA, Tsikriktsis N (2006) Organizational learning curves for customer dissatisfaction: Heterogeneity across airlines. *Management Sci.* 52(3):352–366.

Lapré MA, Nembhard IM (2010) Inside the organizational learning curve. *Foundations Trends Tech., Inform. Oper. Management* 4(1):1–103.

Lu Y, Hechling A, Olivares M (2014) Productivity analysis in services using timing studies. Working paper, Columbia University, New York.

MacDuffie JP (1997) The road to "root cause": Shop-floor problem-solving at three auto assembly plants. *Management Sci.* 43(4):479–502.

Maddala GS (1983) *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge, UK).

Narayanan S, Balasubramanian S, Swaminathan JM (2009) A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Sci.* 55(11):1861–1876.

Phillips R, Şimşek AS, Ryzin Gv (2015) The effectiveness of field price discretion: Empirical evidence from auto lending. *Management Sci.* 61(8):1741–1759.

Pierce L, Snow D, McAfee A (2015) Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Sci.* 61(10):2299–2319.

Pinedo ML (2012) *Scheduling: Theory, Algorithms, and Systems* (Springer, New York).

Pinedo ML, Yen BP-C (1997) On the design and development of object-oriented scheduling systems. *Ann. Oper. Res.* 70(1):359–378.

Pisano GP, Bohmer RMJ, Edmondson AC (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci.* 47(6):752–768.

Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing Service Oper. Management* 14(4):512–528.

Roberti R, Bartolini E, Mingozzi A (2015) The fixed charge transportation problem: An exact algorithm based on a new integer programming formulation. *Management Sci.* 61(6):1275–1291.

Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL (2014) Complexity-augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing Service Oper. Management* 16(3):329–345.

Schultz KL, Juran DC, Boudreau JW (1999) The effects of low inventory on the development of productivity norms. *Management Sci.* 45(12):1664–1678.

Schultz KL, McClain JO, Thomas LJ (2003) Overcoming the dark side of worker flexibility. *J. Oper. Management* 21(1):81–92.

Schultz KL, Schoenherr T, Nembhard D (2010) An example and a proposal concerning the correlation of worker processing times in parallel tasks. *Management Sci.* 56(1):176–191.

Schultz KL, Juran DC, Boudreau JW, McClain JO, Thomas LJ (1998) Modeling and worker motivation in JIT production systems. *Management Sci.* 44(12):1595–1607.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Shumsky RA, Pinker EJ (2003) Gatekeepers and referrals in services. *Management Sci.* 49(7):839–856.

Simon HA (1978) Rationality as process and as product of thought. *Amer. Econom. Rev.* 68(2):1–16.

Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Sci.* 58(6):1141–1159.

Staats BR, Brunner DJ, Upton DM (2011) Lean principles, learning, and knowledge work: Evidence from a software services provider. *J. Oper. Management* 29(5):376–390.

Tan TF, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on server's performance. *Management Sci.* 60(6):1574–1593.

Taylor FW (1911) *The Principles of Scientific Management* (Harper & Brothers, New York).

Truong VA (2015) Optimal advance scheduling. *Management Sci.* 61(7):1584–1597.

Tucker AL (2007) An empirical study of system improvement by frontline employees in hospital units. *Manufacturing Service Oper. Management* 9(4):492–505.

Tucker AL, Nembhard IM, Edmondson AC (2007) Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Sci.* 53(6):894–907.

van Donselaar K, Gaur V, van Woensel T, Broekmeulen R, Fransoo J (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Sci.* 56(5):766–784.

Zhu E, Crainic TG, Gendreau M (2014) Scheduled service network design for freight rail transportation. *Oper. Res.* 62(2):383–400.