Management Science

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

Task Selection and Workload: A Focus on Completing Easy Tasks Hurts Performance

Diwas S. KC, Bradley R. Staats, Maryam Kouchaki, Francesca Gino

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

MANAGEMENT SCIENCE
Vol. 66, No. 10, October 2020, pp. 4397–4416
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Task Selection and Workload: A Focus on Completing Easy Tasks Hurts Performance

Diwas S. KC,[a] Bradley R. Staats,[b] Maryam Kouchaki,[c] Francesca Gino[d]

[a] Goizueta Business School, Emory University, Atlanta, Georgia 30322; [b] Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599; [c] Kellogg School of Management, Northwestern University, Evanston, Illinois 60208; [d] Harvard Business School, Boston, Massachusetts 02163

Contact: diwas.kc@emory.edu, https://orcid.org/0000-0003-4358-2058 (DSKC); bstaats@unc.edu, https://orcid.org/0000-0002-2674-5831 (BRS); m-kouchaki@kellogg.northwestern.edu, https://orcid.org/0000-0003-1292-5134 (MK); fgino@hbs.edu, https://orcid.org/0000-0002-9794-482X (FG)

Received: June 25, 2017
Revised: March 26, 2018; January 6, 2019; May 6, 2019
Accepted: May 18, 2019
Published Online in Articles in Advance: May 20, 2020

https://doi.org/10.1287/mnsc.2019.3419

Abstract. How individuals manage, organize, and complete their tasks is central to operations management. Recent research in operations focuses on how under conditions of increasing workload individuals can decrease their service time, up to a point, to complete work more quickly. As the number of tasks increases, however, workers may also manage their workload by a different process—task selection. Drawing on research on workload, individual discretion, and behavioral decision making, we theorize and then test that under conditions of increased workload, individuals may choose to complete easier tasks to manage their load. We label this behavior task completion preference (TCP). Using six years of data from a hospital emergency department, we find that physicians engage in TCP, with implications for their performance. Specifically, TCP helps physicians manage variance in service times; however, although it initially appears to improve shift-level throughput volume, after adjusting for the complexity of the work completed, TCP is related to worse throughput. Moreover, we find that engaging in easier tasks compared with hard ones is related to lower learning in service times. We then turn to the laboratory to replicate conceptually the short-term task selection effect under increased workload and show that it occurs because of both fatigue and the sense of progress individuals get from task completion. These findings provide another mechanism for the workload-speedup effect from the literature. We also discuss implications for both the research and the practice of operations in building systems to help people succeed.

History: Accepted by Serguei Netessine, operations management.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3419.

Keywords: healthcare • knowledge work • decision making • discretion • workload

In the past the man has been first; in the future the system must be first. This in no sense, however, implies that great men are not needed. On the contrary, the first object of any good system must be that of developing first-class men.
(Taylor 1911, p. 2)[1]

1. Introduction

Since its roots in the scientific management movement that arose in response to the operational complexity of the Second Industrial Revolution, operations management has sought to improve system design to better match supply to demand (Smiddy and Naum 1954). Throughout the 20th century, that meant improving the structure of work, such as better scheduling, superior models for inventory, or enhanced call routing (e.g., Zipkin 2000, Gans et al. 2003, Pinedo 2012). With few exceptions, such as Wickham Skinner's work (Hayes 2002), when people were considered in these models, they were treated as fixed entities to manipulate.

As Taylor's quote above notes, people are different from the inanimate objects, such as inventory or machines, upon which the field of operations often focuses its attention. People can develop, or change, as a result of the system in which they operate. Individuals respond to stimuli and alter their behavior—to improve (or not) performance. The individuals' role in operations grows more central as we see an ongoing increase in service and knowledge operations where humans' ability to learn and adapt often serves as a primary source of competitive advantage.

One of the early findings within the people-centric operations literature was that the rate at which individuals work is not exogenously determined (Schultz et al. 1999). Using multiple settings within healthcare, KC and Terwiesch (2009) show that service rates are endogenous to load. Multiple papers have built on this finding to show how load alters behavior in an operating system (Staats and Gino 2012, Green et al. 2013, Tan and Netessine 2014, Kim et al. 2015, Kuntz et al. 2015, Berry Jaeker and Tucker 2017). A key assumption in this work is that as individuals experience more load, they choose to work faster in the short term, although this speeding up may negatively impact long-term performance.

In this paper we offer a different explanation as to why performance may improve as workload increases: task selection. Recent work in people-centric operations has highlighted that individual discretion—a person's decision about how to alter her work—may have important operational consequences (van Donselaar et al. 2010, Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015, Freeman et al. 2017). In this paper we build on this work considering discretion, to examine whether individuals alter their task selection when workload increases. We propose that people show a task completion preference (TCP) by choosing easier tasks (tasks that can be completed in a shorter amount of time and require less cognitive effort) over difficult ones under states of higher workload compared with states of lower workload. We consider whether this affects both short- and long-term productivity.

To examine our research questions, we rely on data both from the field and from the laboratory. With respect to the former, we investigate an important setting, emergency medicine. We analyze six years of data—more than 230,000 patient encounters—from a major metropolitan hospital. With detailed data, we are able to reconstruct the load of the system, as well as the available patients for pickup, at time of pickup. Thus, as econometricians we see the same information that the doctor saw when she made her patient pickup decision. We then analyze both the short-term and the longer-term impact on operational productivity.

To further study the individual decision-making process, we turn to controlled experiments. Coupling the laboratory with field data permits us to conceptually replicate a key finding and understand the mechanism by which it occurs. People-centric operations studies are increasingly combining these approaches (e.g., Buell et al. 2017, Staats et al. 2018).

In both the laboratory and the field, we find that, on average, people show a task completion preference as load increases. In other words, we find that when the level of workload increases, workers systematically select easier tasks over difficult tasks, exhibiting TCP. This is meaningful because we know that how individuals manage and process their workload has major implications for the performance of workers in the modern workplace. Using the laboratory, we examine the mechanisms by which task completion impacts performance. We find that stress does not mediate the relationship; however, both fatigue and a sense of progress do. In other words, workers may take on simpler tasks when they get tired (an effect our field data support). In addition, under load, it makes workers feel good to complete the tasks, even if the tasks are easy.

However, we find that the strategy of selecting easier tasks may be misleading. In the immediate short term, a single shift, picking easier tasks is associated with a higher throughput volume (total patients seen during the shift); easier tasks are completed quicker than more difficult tasks. Moreover, TCP appears to offer temporary relief during high-workload periods, reducing overall task completion time variation. By picking up easier cases when workload increases, a TCP strategy prevents the overall task completion times from significantly increasing during periods of high workload.

Despite these benefits, further analysis suggests that TCP's impact on performance is more complex. First, TCP may not actually improve shift-level throughput volume. When we move from counting patients processed to a complexity-adjusted measure (relative value units, or RVUs; see Section 3.4 for more details), we find that TCP leads to lower shift-level RVUs. Moreover, when we examine performance over time, we find that greater cumulative volume in hard cases is related to decreased service times and higher RVUs per patient (i.e., learning). Thus, by following a TCP performance strategy, one's long-term productivity may be compromised.

Overall, our study is the first to make several unique contributions to the literature. First, we provide evidence of task completion preference in task selection. In so doing, we provide a new potential explanation for why increased workload might appear to improve immediate performance—people select easier, quicker tasks. Second, we find that this strategy does help manage variation in service times during a shift. Third, we show that if one just considers the number of tasks completed, then TCP improves the throughput volume within a shift. However, when we adjust throughput volume for complexity, and so instead consider a more appropriate measure that captures the actual work being done, we find that selecting easier tasks is related to worse throughput volume performance. Although workers may think they are improving system performance, seeing as easier tasks are completed quicker than more difficult tasks, this is not the case when one accounts for the underlying work completed. Fourth, we find that learning is compromised when individuals pursue a TCP strategy. We show that completing more difficult tasks is related to learning improvement (in service time reductions and RVU increases) compared with completing simpler tasks. Finally, as a part of these models, we show the mechanisms through which completion benefits performance—fatigue and the positive feelings that accrue as work is finished. These contributions have important implications for the design and organization of work and for managing and evaluating worker productivity more broadly.

2. Hypothesis Development

2.1. Literature Review

In this paper we bring together three streams of work. The first is the operations literature on capacity sizing and server scheduling (e.g., Crabill 1972, Stidham and Weber 1989, Pinedo 2012). This literature covers a wide variety of different problems, including allocating capacity (Green et al. 2006), sequencing work (Gerchak et al. 1996), and scheduling service appointments (Bassamboo and Randhawa 2016). KC and Terwiesch (2009) contribute to this literature by identifying that individual service rates that were often treated as fixed and exogenous were, in fact, endogenous and varying to load. In addition, KC and Terwiesch (2009) also show that although service rates initially increase with higher levels of load, they then can decrease when this load is maintained for a long period of time.

The field has built significantly on this paper. For example, Tan and Netessine (2014), drawing on the speed-quality trade-off literature (Hopp et al. 2007, Debo et al. 2008, Anand et al. 2011), hypothesize that as workload increases, individuals alter not only the speed of the service that they offer but also the quality of the service. Using a sample of restaurant servers, the authors find an inverted U-shaped relationship between workload and meal duration as servers adjust service quality to maximize overall revenue. Kuntz et al. (2015) add to the literature by focusing on quality as an outcome. Using hospital data, they find that when workload exceeds a tipping point (92.5% in their data) then in-hospital mortality increases. Also, examining hospital data, Berry Jaeker and Tucker (2017) find that past a certain level of system congestion, patient length of stay increases as the patients left in the system have high demands. Other papers have considered the role of workload, for example, in intensive care admissions' decision (Kim et al. 2015) and emergency department service times (Batt and Terwiesch 2017). We contribute to this literature as we are to our knowledge the first to consider, directly, the role of individual task selection in the workload-speedup effect.

Second, our work builds on literature that studies worker discretion in operating systems. Much of the traditional scheduling and routing literature has taken for granted that once a schedule is set, then it is executed. This is perhaps true for machines but rather less so for humans who have task discretion—the ability to select their next task. Discretion has been examined from several perspectives in operations, including the routing decision (Shumsky and Pinker 2003, Saghafian et al. 2014, Freeman et al. 2017), capacity allocation (Kim et al. 2015), the trade–off between speed and quality (Hopp et al. 2007, Anand et al. 2011, Powell et al. 2012), whether to work in a dedicated versus a pooled queue (Song et al. 2015), and the determination of processing times with different inventory levels (Schultz et al. 1998, 1999).

The literature has found that worker's discretion—whether in the examples in the prior sentence or the workload examples in the previous paragraph—has an impact on operational outcomes. For example, Ibanez et al. (2018) find that radiologists reorganize their work queue, often to choose the shortest task in their queue or to select a task that is similar to the previously completed task, but that in so doing, performance worsens. We contribute to this research by further unpacking the role of discretion in operating systems. Our work is most similar to Ibanez et al. (2018), but whereas they look at how individuals sequence tasks given a preassigned workload, we consider how different workload conditions lead individuals to self-select tasks from a larger set of available tasks, thereby varying the overall work that gets completed. In addition, we consider not only the short-term effects of the choice but also the long-term effects.

Finally, we build on work that examines how individuals deviate from rational agent decision models. Both Bendoly et al. (2006) and Gino and Pisano (2008) provide literature reviews. Suboptimal decision making has been shown in many areas such as forecasting (Kremer et al. 2011), contracting (Davis et al. 2014), and inventory management (Schweitzer and Cachon 2000). Most related to our work is Amar et al. (2011), who show that individuals pay back smaller, lower-interest debts, rather than portions of larger, higher-interest debts, to "complete" the smaller debts. We contribute to this work by suggesting that individuals discount task complexity and focus on task completion under increased workload. We study operational tasks and so both identify the task completion preference and explore mechanisms that may drive the effect. We now motivate our hypotheses.

2.2. Task Selection and Physician Workload

Our first research question asks whether people are more likely to alter their task selection, selecting easier tasks, when they encounter high levels of workload rather than low. Prior to answering this, we make two important points. First, we consider what type of task someone chooses, conditional on choosing a task. In other words, a third option that we do not examine is the decision to take on no additional work. Second, one could study different types of workload: individual or system. We consider the individual's workload, as this is the work for which she is directly accountable.

This is also consistent with prior literature (e.g., KC et al. 2020).

On the one hand, high workload may be related to greater cognitive engagement, leading the worker to enter a state of "flow" (Csikszentmihalyi 1996). Being in a state of engagement is desirable and motivating for the individual worker; workers in a state of flow are more likely to continue to challenge themselves with difficult tasks over the duration of their work period to remain in that state (see also, e.g., Hackman and Oldham 1976 and Grant et al. 2010).

However, on the other hand, there are four theoretical reasons to posit that workers select easier tasks when workload is higher. First, people seek to make progress in their work. Completing work makes people feel good (Amabile and Kramer 2011). Individuals anticipate this positive feeling, and it may motivate them to work harder and seek the experience again (Weick 1984). Moreover, as noted by the goal gradient hypothesis, individual motivation increases as goal completion draws near (Heilizer 1977). Kivetz et al. (2006) find that customers at a coffee shop who are given a free coffee card with 2 out of 12 stamps already punched fill their card quicker than those given a card requiring 10 stamps (with zero punched). Deo and Jain (2019) and Chan (2018) both find that healthcare providers work faster near the end of their shift.

A sense of progress may impact individual's baseline task selection. When choosing between multiple tasks, individuals may choose the easiest to make progress toward their overall goal. Ibanez et al. (2018) show that one reason radiologists change the task order in their queue is to select the shortest task.[2] Here, we are interested in how task selection changes with workload. We hypothesize that a focus on completing tasks is amplified with increasing workload. As people experience higher levels of workload, the desire for progress may increase as a need to alleviate the burden of the load. Higher workload taxes cognitive resources, and so to compensate for this effect, individuals choose easier tasks.

The second reason that individuals may focus on task completion is fatigue. As individuals tire, performance may suffer. For example, Dai et al. (2015) find that as caregivers get further into their shifts, they are less likely to comply with standard processes, such as hand hygiene. In our paper, this suggests that as workload increases and individuals grow fatigued, then they may be more likely to choose easier tasks that can be completed with less effort.

Third, individuals may focus on task completion because of stress. When the amount of work to complete increases, individuals may feel more anxiety or stress. Prior work shows that anxiety changes how individuals view their position—namely, they view the situation as more of a threat (Staw et al. 1981). With this change in perspective, individuals restrict their information processing and use simpler rules to decide which task to select next. As a result, they may select easier tasks to offset the increasing stress from workload.

Finally, individuals may choose to prioritize task completion for operational reasons. One common heuristic taught to practitioners is that to minimize mean throughput time, one should complete the shortest task first (Cachon and Terwiesch 2009). Though an individual may choose to follow this rule at any point, as workload increases, overall flow times grow more salient and so individuals may default to this rule. As a result of these factors, we hypothesize the following.

Hypothesis 1. *People are more likely to select easy tasks than harder tasks when they are confronted with a high versus low workload.*

2.3. Mechanism of Task Completion Preference

As our second hypothesis, we turn to the potential mechanism that may drive the effect. Above, we highlighted four possible mechanisms: (1) sense of progress, (2) fatigue, (3) stress, and (4) operational concerns. Here, we focus on the three psychological mechanisms. Prior work talks about all three and does not specify which may drive the effect. It is plausible that multiple could, or alternatively, one could dominate. Therefore, we offer each as an independent hypothesis.

Hypothesis 2A. *People are more likely to select easy tasks than harder tasks when facing a high versus a low workload because it gives them a sense of progress.*

Hypothesis 2B. *People are more likely to select easy tasks than harder tasks when facing a high versus a low workload because of fatigue.*

Hypothesis 2C. *People are more likely to select easy tasks than harder tasks when facing a high versus a low workload as a result of stress.*

2.4. Performance Consequences of Task Completion Preference

We now turn to the performance consequences of selecting easier tasks. To answer the question in this context, it is necessary to first specify the performance metric. To permit us to account for the fact that a physician is managing multiple cases simultaneously and may trade off his or her focus across these cases (KC 2013), we consider shift-level throughput volume—the number of cases completed during the shift. Not only is this measure operationally appropriate, but it allows us to consider the short-term effect of a choice as a shift captures no more than a day (a length of time used previously in the literature to capture the short term; e.g., Staats and Gino 2012).

How then will selecting easier tasks when workload increases relate to shift-level throughput (in terms of task completion)? Whether it is done for psychological or operational reasons, the expectation is that choosing the easier task will result in quicker service times, helping to manage the workload. As noted above, completing tasks offers motivational benefits that may translate to a general speeding-up effect and thus a completion of more work. Moreover, if one is tired and stressed, then selecting easier tasks may create an opportunity to keep going where a difficult task might stymy progress. Eventually, this is an empirical question, but given this logic, we hypothesize the following.

Hypothesis 3. *Selecting easy tasks over hard tasks when workload increases improves shift-level throughput volume.*

Selecting easier tasks may also offer shift-level benefits in managing variability. Variability in service processing rates can negatively impact the performance of queuing systems (e.g., see Cachon and Terwiesch 2009). As such, variability reduction through strategies such as demand smoothing, processing time standardization, and product variety limits are widely studied in operations management. TCP may also be such a strategy.

Under conditions of increasing workload, a worker is spread thin, dividing limited time and cognitive resource across competing activities. Also, the dependence on other constrained resources (e.g., physical equipment, space, and personnel) may worsen service rates. We know from basic queuing theory that as utilization increases, service time increases dramatically. This means that when a difficult case (which already takes longer) is selected, the effects of queuing are further compounded. Therefore, by selecting an easy case next, the worker can prevent the disproportionate increase in the service time. In other words, TCP uses the convex service-time–utilization relationship to prevent dramatic increases in service time. Thus, the fact that task completion preference leads to a more uniform service time distribution is an interesting, and potentially unintended, positive side effect. Finally, workers who are fatigued may take longer to complete assigned tasks. As such, taking on easier tasks during periods of high workload would constitute a form of demand smoothing, where the worker matches periods of low processing capacity availability with less onerous tasks. Overall, we postulate that TCP as an adaptive behavior may smooth demand, and so we hypothesize the following.

Hypothesis 4. *As workers select a greater number of easier tasks when they have higher workload, they have lower variability in processing times during a shift.*

Selecting an easy task over a hard task may improve the short-term throughput volume and help manage variability with increased workload, but it also may lead to longer-term consequences. Here, we consider the implications of a workload-driven choice between easier and harder tasks. In other words, over extended periods when individuals may have gained easy or hard experience, some of which is attributed to TCP, how does it affect performance?

Experience and its effect on performance and choices has received significant attention in the operations literature (Lapré and Nembhard 2010, Bolton et al. 2012, KC and Staats 2012, Arlotto et al. 2014). Usually, experience is beneficial, as people accrue experience they learn and improve, although experience can lead to suboptimal choices (Staats et al. 2018). Recent work on experience highlights that not all experience is created equal, and some experience may be more beneficial for performance and learning (Huckman et al. 2009, Narayanan et al. 2009, Staats and Gino 2012).

The underlying arguments in this work suggest that harder tasks might hold more learning content than easier ones. Learning from experience is premised on a learning curve—with repeated experience, an individual improves at a task (Lapré and Nembhard 2010). Some improvement comes from learning a routine, but much of it comes from learning the intricacies of the task. Moreover, theoretical research on learning (Zangwill and Kantor 1998) suggests that overall learning curves are made up of many smaller learning curves. Thinking in this way helps one see that learning curves for harder tasks involve many subtasks. By focusing on hard tasks, an individual may benefit from learning opportunities, gaining new skills or challenging herself to think deeply on a topic. Interestingly, decision-making research suggests that if individuals focus on subgoals, such as executing easy tasks to keep the system moving, then they may keep their focus there and not reallocate their attention to the broader goals—either harder tasks or learning, more generally (Heath et al. 1999, Amar et al. 2011). Thus, we hypothesize the following.

Hypothesis 5. *Over time, cumulative experience with hard tasks will improve service time more than cumulative experience with easier tasks.*

We now turn our attention to testing our hypotheses. We first go to the field to examine whether individuals exhibit a task completion preference by choosing easy tasks over hard tasks under increased workload (Hypotheses 1 and 2B) and then the short-term throughput volume and variability (Hypotheses 3 and 4) and long-term productivity (Hypothesis 5) implications. We then go into the laboratory to replicate conceptually the TCP finding (Hypothesis 1) and investigate the potential mechanism (Hypotheses 2A–2C).

3. Study 1: Task Selection and a Focus on Completion in the Field

In this study we investigate whether task completion preference is exhibited by knowledge workers in the workplace. We specifically look at the task selection behavior of physicians in the emergency department (ED) of a hospital and examine the opera tional implications.

The field setting provided by the ED offers several key features that enable our study. First, the arrival of patients to the ED is an inherently random occurrence. The random arrival of patients to the ED means that a mix of patients available for pickup and processing is exogenous to the existing workload of the ED and of individual physicians. Second, random patient arrivals and a queue of tasks waiting to be processed mean that we observe varying levels of *offered* physician workload, which we can use to identify the task selection effect. Third, the time taken by the physician to process a patient, which is the time elapsed between picking up and discharging the patient, is clearly defined. Finally, we are able to track the performance of individual physicians over long periods of time to assess learning from experience.

We posit that processing time can be broken into two sets of underlying drivers, patient-specific factors and physician-specific factors. In the ED context, easy tasks correspond to treating patients with lower acuity levels. We postulate that under increased levels of workload, physicians are more likely to pick up low-acuity, rather than high-acuity, patients, consistent with a task completion preference. We also postulate that in the short term (defined at the level of a physician-shift), the patient throughput volume increases because of TCP, corresponding to short-term productivity gains. In the long term, we hypothesize that taking on easy patients is associated with lower productivity, as measured by the time taken to treat and discharge patients.

3.1. Setting

The field study context is the medium-sized ED of an East Coast metropolitan hospital that treats a sizable volume of patients each year. Patients arrive unscheduled to the ED. Upon arrival, a patient is seen by the triage nurse, who evaluates the patient's condition, determines the triage acuity level based on an Emergency Severity Index Score (ESI scores range from 1 to 5, with 1 being the most severe), notes the chief complaint for the patient, and creates an electronic record and a physical folder for the patient. The patient's electronic record is then placed in a virtual queue to be processed by a physician, who is specialized in emergency medicine. At any point in time, there are a number of physicians in the ED, each of whom uses a computer terminal for monitoring the queue, picking up new patients, and providing updates to the clinical record upon treatment. Physicians continuously monitor the queue to pick up new patients when not attending to existing patients.

The service process for a patient begins once she is picked up by a physician. The patient service process involves several key events, including evaluating the patient's medical record, physically examining the patient, gathering medical information, ordering tests and procedures, and discharging the patient from the ED. At any point, a physician is often responsible for more than one patient in the ED. The number of patients that a physician concurrently manages may vary significantly over the course of the shift. This change in physician workload is used to explain the physician's choice of whether to pick up an easy or a difficult patient.

3.2. Data Description

We assembled our data from the emergency department for fiscal years 2005–2010 involving over 233,000 distinct patient encounters treated by 84 providers. Our data include patient-level visit information, as well as physician-level and ED-level factors. Most important, we observe the unique patient-physician pair for each patient in the ED, meaning we know the specific physician who treated an individual patient. This information thus allows us to examine the service encounter of each patient, as well as the productivity of each physician.

With our data set we can estimate productivity. We observe the time ($t_{\mathrm{pickup},i,j}$) a patient j is picked up by physician i and when she is discharged ($t_{\mathrm{discharge},i,j}$). This is the time that a patient is under the care of a given physician (physician assignment period) and is simply the difference between the instant the patient is assigned and the instant that she is discharged from the care of the ED physician. This time in service is denoted as the *service time* ($SvcTime_{ijt}$) for patient j treated by physician i beginning at time t, where $SvcTime_{ijt} = t_{\mathrm{discharge},i,j} - t_{\mathrm{pickup},i,j}$.

For each patient who presents in the ED, we observe several clinical variables, including the patient's acuity level, which is used to categorize the patient as easy or difficult. There are five ESI levels of acuity (1–5). To facilitate the analysis and to offer a comparison with Study 2, we dichotomize the acuity score into two groups: [1, 2] = difficult and [3, 4, 5] = easy. This definition of easy versus difficult is based on discussions with ED physicians and triage nurses, who suggested that a cutoff above and below ESI level of 3 represents the most natural binary demarcation. As additional tests of robustness (see the online appendix), we consider a cutoff at ESI level of 4, as well as an alternative definition of easy versus difficult based on the chief complaint of the patient. We find that 50.3% of the patients are low-acuity patients, and the average service time is 3.85 hours (with a standard deviation of 1.85 hours). Also, we observe various sources of patient-level heterogeneity, including age; gender; race; payment status; means of arrival; and temporal variables including time of treatment, day of week, and month of year. The average patient age is 36, but there is significant variation. Approximately half (48%) of the patients are female.

For each patient visit, the chief complaint is recorded. However, this variable is recorded as a free-text field and therefore unusable in its raw form. We utilize the Clinical Classification Software[3] (the CCS score), which codifies the patient's primary medical concern into one of 250+ distinct clinical classifications. We also observe the RVU, or relative value unit, associated with a given patient. RVUs are designed to capture the amount and difficulty of work completed (which is the basis for using RVUs for reimbursement). Specifically, the RVU is a measure of the physician's work associated with the care of a given patient, and it takes into account the physician's time, as well as clinical and technical judgment, effort, and skill, with physician work being the largest component of RVUs.[4] RVUs are used by payers, including Medicare, to reimburse physicians for services rendered. As such, RVUs allow us to examine the implications of TCP on complexity-adjusted productivity. The average RVU is 5.82 units per encounter.

Our operational data allow us to construct a time series for the number of patients under the care of a given physician at a specific point in time (denoted as $PhyLoad_{it}$). Specifically, once a patient is picked up by physician i at time t, the level of workload for that physician is increased by 1. Similarly, at the instant a patient is discharged, the level of workload decreases by 1. In other words, the physician's workload remains constant until either a pickup or discharge event occurs. The physician load at time $t+\Delta t$ is obtained after accounting for the number of patients picked up ($pickups_{t+\Delta t}$) and discharged ($discharges_{t+\Delta t}$) by physician i between t and $t+\Delta t$ as $PhyLoad_{i,t+\Delta t} = PhyLoad_{i,t} + pickups_{i,t+\Delta t} - discharges_{i,t+\Delta t}$. To account for the mix of easy and difficult cases in a physician's workload, we create a measure called $PhyCasemix_{it}$, which is defined as the number of easy cases divided by the total number of patients under the care of physician i at time t; $PhyCasemix_{it}$ is therefore the fraction of easy cases that comprise the physician's workload. In addition to the resources of the attending physician, in-process patients (those who already picked up by *any* physician) also consume the common resources of the hospital such as nurses, room and hallway bed space, and centralized laboratory testing facilities and clinical specialists. To control for the workload on the system, we define a variable, $System\ Load_t$, which is the collective workload that is being processed by physicians. As with the individual physician workload, system load is incremented by 1 each time *any* physician picks up a patient and is reduced by 1 each time *any* patient is discharged from the ED. Specifically, the system load at time $t+\Delta t$ is obtained after accounting for the number of patients picked up ($pickups_{t+\Delta t}$) and discharged ($discharges_{t+\Delta t}$) by in Δt as $System\ Load_{t+\Delta t} = System\ Load_t + pickups_{t+\Delta t} - discharges_{t+\Delta t}$. In addition to the patients that have already been assigned to physicians, there is a queue of patients who have arrived and need to be picked up. We define this volume of patients who are waiting as the waiting load ($WaitLoad_t$). Similarly, the case mix of waiting patients ($WaitCasemix_t$) is defined as the number of easy patients divided by the total number of patients who have yet to be picked up at time t. Finally, our data include the number of ED physicians ($NumPhysicians_t$) working in the ED at time t.

Table 1 provides summary statistics. We find that the system load in the ED is 23.95. However, there is significant variability as indicated by the high standard deviation. The average physician workload is 5.1 patients, and the standard deviation is 3.4 patients.

Our primary objective is to examine physician pickup behavior as a function of physician workload. Specifically, at any given point in time, there is a set of patients in the waiting area, with both high and low levels of acuity. Our data allow us to consider the set of patients available for physician pickup at any point in time. At various times, individual physicians have different levels and case mixes of workload, and this may impact the physician's decision to pick up an easy or a more acute patient. By combining the system- and physician-level factors with the patient clinical considerations, we have a comprehensive data set that allows us to analyze the service encounter, selection activity, and productivity for every single physician in the ED.

3.3. Empirical Specifications

The empirical specifications that we develop below allow us to estimate the pickup behavior, the resulting short-term and long-term productivity, and variability of outcomes of ED physicians.

3.3.1. Physician Task Selection. This subsection presents our empirical model of physician choice. Our approach uses the physician choice of patient to pick up (i.e., revealed preference) as well as physician, patient, and emergency department characteristics to examine the drivers of physician behavior. At any given time t, physician i can choose to pick up either an easy patient ($EASY_j = 1$) or a difficult patient ($EASY_j = 0$) from the set of patients available in the waiting area.

We examine the physician's choice using a logit model. In particular, we are interested in the effects of the physician's workload on the choice of type of

Table 1. Summary Statistics

	Mean	Med.	sd	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. *EASY*	0.50	1.00	0.50	1.00														
2. *SvcTime*	3.85	3.63	1.85	−0.03	1.00													
3. *RVU*	5.82	2.91	7.25	−0.10	0.09	1.00												
4. *Female*	0.47	0.00	0.50	−0.01	0.05	0.02	1.00											
5. *Age*	36.02	34.00	21.02	−0.03	0.12	−0.03	0.02	1.00										
6. *SystemLoad*	23.95	24.00	10.18	−0.15	0.11	0.01	0.04	−0.06	1.00									
7. *PhyLoad*	5.14	4.00	3.42	0.07	−0.08	−0.10	−0.01	0.03	0.29	1.00								
8. *Fatigue*	2.41	1.00	3.82	0.05	−0.15	−0.07	−0.02	0.01	0.07	0.39	1.00							
9. *FatigueRVU*	5.37	0.00	12.34	0.01	−0.11	−0.02	−0.01	−0.01	0.05	0.27	0.74	1.00						
10. *PhyCasemix*	0.49	0.59	0.33	0.65	0.04	−0.07	−0.02	0.01	−0.26	−0.02	0.00	−0.02	1.00					
11. *NumPhysicians*	2.54	2.00	1.50	−0.15	0.07	0.03	0.03	0.00	0.51	−0.02	0.00+	0.001+	−0.17	1.00				
12. *WaitLoad*	13.68	13.00	6.12	−0.21	0.05	0.01	0.04	−0.02	0.79	0.24	0.06	0.03	−0.30	0.60	1.00			
13. *WaitCasemix*	0.51	0.60	0.35	0.69	0.03	−0.07	−0.02	0.00	−0.25	−0.01	0.01+	−0.01	0.95	−0.19	−0.31	1.00		
14. *ExpDiff* ('000s)	0.95	0.57	0.99	−0.42	−0.06	0.05	0.02	0.02	0.11	0.04	−−0.02	0.008+	−0.63	0.05	0.16	−0.60	1.00	
15. *ExpEasy* ('000s)	1.81	1.54	1.44	−0.23	−0.02	0.02	0.01	0.06	0.05	0.05	0.00	0.00	−0.33	−0.02	0.07	−0.32	0.79	1.00

Notes. N = 233,880. All the correlations are statistically significant at the 5% level unless indicated otherwise (+).

patient selected. Under the logit specification, the probability that physician i chooses patient j at time t is given by

$$\Pr_{it}(EASY_j) = \frac{e^{\alpha_i+\tau_t+\mathbf{X}_j\beta+\mathbf{Z}_t\gamma+\theta PhyLoad_{it}+\delta PhyCasemix_{it}+\Phi Fatigue_{it}}}{\sum_{ijt} e^{\alpha_i+\tau_t+\mathbf{X}_j\beta+\mathbf{Z}_t\gamma+\theta PhyLoad_{it}+\delta PhyCasemix_{it}+\Phi Fatigue_{it}}}, \tag{1}$$

where α_i is the fixed effect for physician i, and our identification is therefore driven by intraphysician variation over time. Let $\boldsymbol{\tau}_t$ be a vector of temporal controls that account for seasonality, including the hour of day, day of week, month of year, and year. Let $\mathbf{X}_j$ be a vector of patient-level controls observable by the physician prior to patient pickup. Given the diagnostic nature of the ED visit, detailed information on the patient's condition is determined after the patient is evaluated by the physician and is not available to the physician prior to pick up. As such, $\mathbf{X}_j$ includes the patient's demographic factors including age, gender, race, and CCS score (proxy for chief complaint). Let $\mathbf{Z}_t$ be a vector of ED-level factors that change over time; this list includes the number of other physicians in the ED at time t, the number of patients currently under the care of *any* physician at time t (system load), the number of patients who have not yet been picked up from the waiting area (wait load), and the fraction of patients in the waiting area that are considered easy cases (wait case mix); it also includes the $ShiftEnd_t$, defined as the time left in the physician's shift.

$PhyLoad_{it}$ is the number of patients under the care of physician i at time t. $PhyCasemix_{it}$ is the fraction of easy patients in the physician's workload, and it allows us to adjust for the intensity of the physician's case mix. Finally, to examine whether fatigue affects the patient pickup decision, we construct $Fatigue_{it}$, which is based on the amount of work that the physician has already completed during the shift. Specifically, $Fatigue_{it}$ is operationalized as the number of patients that the physician has finished (i.e., discharged) at time t since the start of the shift. As another robustness test, we consider $FatigueRVU_{it}$, the RVU units completed by physician i at time t. By construction, and consistent with the patient flow literature, these workload variables, including $PhyLoad_{it}$, $Fatigue_{it}$, $WaitLoad_t$, and $SystemLoad_t$ as well as the patient case mixes, are realized prior to the pickup of patient j. Empirical specification (1) corresponds to whether an easy or a difficult patient is picked up. The baseline case is that a difficult case is selected. A positive value for θ thus supports Hypothesis 1, that a physician is more likely to pick up an easy patient as workload increases. Similarly, a positive value for ϕ supports the hypothesis that fatigue leads physicians to pick up easy cases. We use a full information maximum likelihood estimator to obtain the parameters of interest of the choice model given by (1).

3.3.2. Short-Term Productivity Effects. We next examine whether picking up an easy patient is associated with a short-term productivity improvement. Our unit of analysis is the individual physician shift, and we consider the total work completed during a shift as our measure of short-term productivity. There are two reasons for examining productivity at the shift level. First, it allows us to remain agnostic to how the physician is managing her workload over the duration of the shift, because we are only concerned with the aggregate shift productivity resulting from the pickup of easy cases during the shift. Second, shift-level throughput volume is a concrete and tangible outcome measure of short-term productivity, and it provides a useful metric for the physician to assess their performance. In our analyses of shift-level performance, we consider two outcomes: (1) the volume of patients and (2) the total RVUs generated during the shift. We first examine whether taking on easy patients allows the physician to generate higher patient volume per shift (s), using the following specification:

$$\begin{aligned} Vol_{is} = {} & \alpha_i + \boldsymbol{\tau_s} + \beta\overline{SystemLoad}_s + \omega\overline{WaitLoad}_s \\ & + \gamma\overline{PhyLoad}_s + \theta\overline{EASY}_s \\ & + \vartheta\overline{PhyLoad}_s\, X\, \overline{EASY}_s + \varepsilon_{is}. \end{aligned} \tag{2}$$

Here, Vol_{is} is the total volume of patients discharged by physician i during shift s; α_i is the physician fixed effect, which subsumes physician heterogeneity such as ability, motivation, and practice style; and $\boldsymbol{\tau_s}$ captures temporal factors associated with the shift, including the year, month, day of week, starting hour of the shift, and shift duration. Because system load, waiting load, and the physician's workload vary during the course of the shift, we use their shift-level averaged values. Similarly, we compute the share of easy patients picked up during the shift (or the fraction of easy patients, relative to the total number of patients), denoted as $\overline{EASY}_s$. Hypothesis 3 posits that we expect to see more patient throughput volume as a result of the pickup of easy cases during the shift. A value of $\vartheta > 0$ would confirm the hypothesis that by picking up easy patients when the load is high, physicians are able to generate a higher shift-level throughput volume.

We similarly consider if TCP leads to more work completed, as indicated by the total RVUs produced (sum of the RVUs for each patient discharged during a shift). Note the expected result from this model does not trivially follow from the prior one. Easier tasks are completed faster (good for productivity) but involve lower RVUs (bad for productivity). We modify the specification above by replacing physician shift volume with total RVUs generated in the shift:

$$\begin{aligned} RVU_{is} = {} & \alpha_i + \tau_s + \beta\overline{SystemLoad}_s + \omega\overline{WaitLoad}_s \\ & + \gamma\overline{PhyLoad}_s + \theta\overline{EASY}_s + \vartheta\overline{PhyLoad}_s\, X\, \overline{EASY}_s \\ & + \varepsilon_{is}. \end{aligned} \tag{3}$$

A value of $\vartheta > 0$ indicates that picking up easy cases during high workload leads to more units of work completed, as provided by the amount of RVUs.

In addition to impacting the time taken to complete a given amount of work, TCP may also change the variability in task completion times. We next consider the overall variability of the service times of the patients over the course of the shift as a result of TCP. To quantify this effect, we first evaluate the coefficient of variation in the service time for patients during a shift as our measure of shift-level service time variability. We then assess whether shift-level variability in service times is affected by TCP using the following empirical specification at the shift level:

$$\begin{aligned} \mathrm{CV}(SvcTime)_{is} = {} & \alpha_i + \tau_s + \beta\overline{SystemLoad}_s \\ & + \omega\overline{WaitLoad}_s + \gamma\overline{PhyLoad}_s + \theta\overline{EASY}_s \\ & + \vartheta\overline{PhyLoad}_s\, X\, \overline{EASY}_s + \varepsilon_{is}. \end{aligned} \tag{4}$$

In the specification above, $\mathrm{CV}(SvcTime)_{is}$ is the coefficient of variation in service time for physician i during shift s. We control for temporal factors such as seasonality, as well as the average system load, wait load, and physician workload during the shift. A negative value for ϑ provides support for the hypothesis that TCP leads to a reduction in the variability of service processing times for the physician.

3.4. Results: Task Selection and Short-Term Productivity

We find that physicians are more likely to pick up easier patients (compared with picking up a difficult patient) when the workload is higher. In Table 2, specification (1) excludes the time fixed effects, and we find that coefficient of physician workload is 0.075 ($p < 0.01$). Specification (2) excludes the patient-level controls. We find that excluding the patient-level heterogeneity still yields the same general result (0.11, $p < 0.01$). This suggests that temporal factors such as the time in shift do not substantially drive observed short-term productivity effects. Specification (3) includes an alternative measure of fatigue, based on the RVUs completed. We again find that in increase in fatigue leads to a greater likelihood of picking up an easy patient (0.002, $p < 0.01$). Specification (4) is our full model and accounts for physician

Table 2. Task Selection Preference

	(1)	(2)	(3)	(4)
Physician fixed effects	X	X	X	X
Time fixed effects		X	X	X
Patient controls	X		X	X
Female	0.051***		0.054***	0.051***
	(0.014)		(0.013)	(0.015)
Age	−0.003***		−0.004***	−0.003***
	(0.000)		(0.000)	(0.000)
Time Left in Shift	0.001***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
PhyCasemix	−0.484***	−0.607***	−0.757***	−0.531***
	(0.077)	(0.076)	(0.089)	(0.083)
NumPeers	0.002	−0.009	0.007	−0.001
	(0.006)	(0.006)	(0.007)	(0.006)
WaitLoad	−0.010***	−0.010***	−0.007**	−0.009***
	(0.002)	(0.003)	(0.003)	(0.002)
WaitCasemix	6.555***	6.436***	6.576***	6.533***
	(0.068)	(0.064)	(0.073)	(0.068)
SystemLoad	−0.006***	−0.009***	−0.007***	−0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
PhyLoad	0.075***	0.111***	0.089***	0.079***
	(0.004)	(0.006)	(0.005)	(0.004)
Fatigue	0.050***	0.079***		0.053***
	(0.004)	(0.004)		(0.004)
FatigueRVU			0.002***	
			(0.001)	
Number of observations	208,105	216,558	156,565	208,105
Log likelihood	−76,241.47	−83,841.47	−58,415.43	−75,933.80

Notes. Categorical patient controls include CCS and race. Standard errors clustered by physician are in parentheses.
***$p < 0.01$; **$p < 0.05$.

fixed effects, temporal considerations, system-level factors, and patient heterogeneity. We find that our estimate for workload is 0.079 ($p < 0.01$). This corresponds to increased odds of 8% in picking up an easier patient when the physician's workload marginally increases by 1. We find that fatigue also influences the likelihood of an easy patient pickup. From specification (4), we find that an increase in fatigue based on the discharge of one additional patient is associated with a 5% increase in the odds of picking up an easy patient (coefficient = 0.053, $p < 0.01$).

As additional tests of robustness, we consider cutoffs for easy versus difficult at an alternative ESI level of 4. We also define easy versus difficult based on the medical condition (using the CCS index). Finally, we consider a linear probability model and a multinomial logit model to study whether the functional specification impacts our results (see the online appendix). We find that our results are robust to various alternative definitions of easy and difficult and to alternative model specifications.

We next examine the effect of TCP on shift-level productivity, as measured by the number of patients discharged (see Table 3). Specification (1) excludes the physician fixed effect but includes temporal sources of heterogeneity as well as system-level factors. We find that the interaction between the physician workload and the share of easy patients in the shift is positive (0.140, $p < 0.05$). In other words, picking up a larger share of easy patients during busier shifts increases physician throughput. Specification (2) includes the physician fixed effect but excludes the temporal sources of heterogeneity. We find that the effect of TCP on short-term productivity continues to be positive (0.112, $p < 0.05$); this suggests that physician-level factors or seasonality is unlikely to confound our estimates. Specification (3) excludes the interaction effect; we find that as the share of easy patients increases, overall shift throughput increases, holding workload constant (0.694, $p < 0.01$). Specification (4) is our full model, which includes physician fixed effects, temporal controls, and system-level factors. We find that the interaction term between physician workload and the fraction of easy cases is 0.147 ($p < 0.01$). Collectively, these results demonstrate that a higher share of easy patients increases overall volume. Moreover, the increase is greater during periods of higher physician workload. We find that our estimates are

Table 3. Short-Term Productivity on Shift Throughput

	(1)	(2)	(3)	(4)
Physician fixed effects		X	X	X
Time fixed effects	X		X	X
SystemLoad	−0.172***	−0.127***	−0.122***	−0.123***
	(0.011)	(0.008)	(0.008)	(0.008)
WaitLoad		0.235***	0.229***	0.232***
		(0.016)	(0.015)	(0.015)
PhyLoad	2.657***	2.680***	2.718***	2.641***
	(0.033)	(0.032)	(0.021)	(0.033)
Fraction of Easy Patients	0.183	−0.205	0.694***	0.143
	(0.149)	(0.162)	(0.100)	(0.147)
PhyLoad × *Easy Patient Fraction*	0.140**	0.112**		0.147***
	(0.056)	(0.054)		(0.056)
Number of observations	21,279	21,279	21,279	21,279
Log likelihood		−47,333.94	−47,261.14	47,234.25

Note. Standard errors clustered by physician are in parentheses.
*** $p < 0.01$; ** $p < 0.05$.

not significantly different across the model specifications, indicating robustness of our results. Combined with our task selection results, Table 3 confirms our hypothesis that by selecting easier tasks, short-term throughput volume improves.

We next examine the effect of TCP on the number of RVUs generated during the shift (Table 4). As discussed earlier, RVUs capture the amount of work completed over the course of the shift. Specification (1) excludes the physician fixed effect but includes the temporal and system-level heterogeneity. We find that the interaction effect between the physician workload and the share of easy patients is negative (−4.64, $p < 0.001$), indicating that taking on easier patients during busy periods leads to lower overall RVUs being generated. Specification (2) includes the physician fixed effect but excludes the temporal sources of seasonality. We find the coefficient of interest is relatively unchanged (−4.50, $p < 0.01$), suggesting that our estimate is not confounded by individual physician heterogeneity or the temporal fixed effects. Specification (3) examines easy patients and overall RVU throughput. We see that as the share of easy patients increases, overall RVUs decrease (−6.67, $p < 0.01$). Specification (4) is our full model and includes the physician, temporal, and system-level factors. We find that the interaction between physician load and the share of easy patients is statistically significant (−4.49, $p < 0.01$). The results from Table 4 indicate that in the short term, TCP leads to fewer RVU units being generated.

Finally, we consider the effect of TCP on service time variability (Table 5). Across all the model specifications, we find that the interaction term between physician workload and the fraction of easy patients is negative and statistically significant. In particular, the

Table 4. Short-Term Productivity on Shift RVUs

	(1)	(2)	(3)	(4)
Physician fixed effects		X	X	X
Time fixed effects	X		X	X
SystemLoad	−0.262***	−0.453***	−0.291***	−0.254***
	(0.037)	(0.045)	(0.038)	(0.037)
WaitLoad	0.533***		0.547***	0.462***
	(0.084)		(0.084)	(0.079)
PhyLoad	11.560***	11.779***	9.312***	11.663***
	(0.355)	(0.345)	(0.210)	(0.350)
Fraction of Easy Patients	11.255***	8.251***	−6.676***	10.240***
	(1.489)	(1.425)	(1.105)	(1.452)
PhyLoad × *Easy Patient Fraction*	−4.643***	−4.500***		−4.488***
	(0.367)	(0.357)		(0.364)
Number of observations	21,004	21,004	21,004	21,004
Log likelihood		−96,656.99	−96,779.88	−96,570.98

Note. Standard errors clustered by physician are in parentheses.
*** $p < 0.01$.

Table 5. Short-Term Variability in Service Time on Shift

	(1)	(2)	(3)	(4)
Physician fixed effects		X	X	X
Time fixed effects	X		X	X
SystemLoad	−0.636***	−0.759***	−0.636***	−0.633***
	(0.040)	(0.052)	(0.040)	(0.039)
WaitLoad	0.704***		0.694***	0.686***
	(0.072)		(0.073)	(0.071)
PhyLoad	0.593***	0.596***	0.402***	0.605***
	(0.105)	(0.107)	(0.088)	(0.107)
Fraction of Easy Patients	1.710**	1.761**	−0.179	1.485*
	(0.836)	(0.813)	(0.483)	(0.824)
PhyLoad × *Easy Patient Fraction*	−0.436***	−0.401***		−0.398***
	(0.131)	(0.131)		(0.132)
Number of observations	21,279	21,279	21,279	21,279
Log likelihood		−88,530.76	−88,523.28	−88,519.04

Note. Standard errors clustered by physician are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

exclusion of physician fixed effects (specification (1), estimate −0.436, $p < 0.01$) and the exclusion of seasonal heterogeneity (specification (2), estimate = −0.401, $p < 0.01$) do not significantly deviate from the full model specification (specification (4), estimate = −0.398, $p < 0.01$). This suggests that unobserved physician-level and seasonal factors are unlikely to confound our results. Collectively, the results from specifications (1)–(4) show that taking on easier patients is particularly helpful during busy periods in reducing service time variability, supporting Hypothesis 4.

3.5. Long-Term Productivity Effects

To examine the long-term effects of task selection, we construct a learning-curve model that links the cumulative volume of patients treated to the service time on subsequent patients. We seek to attribute the separate effects of experience obtained through easy and difficult tasks on subsequent productivity. Our estimation strategy takes the form of a panel, where we follow individual physicians over time, and we link the cumulative experiences to future performance. Our empirical specification takes the following form:

$$\log(SvcTime_{ijt}) = \alpha_i + \boldsymbol{\tau_t} + \boldsymbol{X_j}\beta + +\boldsymbol{Z_t}\gamma + \boldsymbol{S}_{it}\rho + \rho_1 Exp_{it} + \rho_2 FracDiffExp_{it} + \varepsilon_{ijt}. \quad (5)$$

In the specification above, the outcome of interest is the service time on patient j seen by physician i at time t. Let α_i be the physician fixed effect that captures the baseline productivity of the physician. Let $\boldsymbol{\tau_t}$ be a set of temporal factors, including the year, month, day of week, and hour of treatment start. Let $\boldsymbol{X_j}$ be a vector of patient-level controls, including the ESI severity score; CCS index; and demographic factors such as age, gender, and race for patient j. Let $\boldsymbol{Z_t}$ be a vector of system-level factors specific to the ED, including the waiting load, system load, and the number of physicians working the ED at time t; let $\boldsymbol{S}_{it}$ be a vector of physician-level variables including the workload and the workload case mix at time t. Let Exp_{it} be the cumulative volumes of cases processed by physician i since the start of the study period: $FracDiffExp_{it} = \frac{ExpDiff_{it}}{ExpDiff_{it}+ExpEasy_{it}}$. In other words, the fraction of difficult experiences is simply the ratio of the cumulative difficult experiences by time t for physician i ($ExpDiff_{it}$) to the total cumulative difficult and easy experiences ($ExpDiff_{it} + ExpEasy_{it}$). In specification (5), a negative value for ρ_2 would indicate that with greater experience on difficult tasks relative to the experience on easy tasks, physicians become faster at processing patients. We also consider an alternative performance measure, given by the number of RVUs generated per patient. We posit that learning through experience enables physicians to become more productive, as measured by RVUs generated. To assess the effect of experience on RVU production, we modify the above specification but replace service time with the RVUs generated for each patient j as follows:

$$RVU_{ijt} = \alpha_i + \boldsymbol{\tau_t} + \boldsymbol{X_j}\beta + \boldsymbol{Z_t}\gamma + \boldsymbol{S}_{it}\rho + \rho_1 Exp_{it} + \rho_2 FracDiffExp_{it} + \varepsilon_{ijt}. \quad (6)$$

In this specification, values of $\rho_2 > 0$ would support our hypothesis that experience of difficult cases lead physicians to be more productive compared with the experiences of easy cases.

3.6. Results: TCP and Long-Term Learning

Table 6 presents the results for the effect of TCP on long-term productivity as measured by the service time. Specification (1) excludes the physician fixed effect. The explanatory variable for total experience has been standardized, and we find that as the

Table 6. Learning Effects on Service Time

	(1)	(2)	(3)
Physician fixed effects		X	X
Time fixed effects	X		X
Female	0.022***	0.022***	0.022***
	(0.002)	(0.003)	(0.003)
Age	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
SystemLoad	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)
PhyLoad	−0.005***	−0.005***	−0.005***
	(0.000)	(0.001)	(0.001)
PhyCasemix	0.125***	0.122***	0.069***
	(0.011)	(0.016)	(0.016)
WaitLoad	−0.005***	−0.004***	−0.003***
	(0.000)	(0.001)	(0.001)
WaitCasemix	−0.055***	−0.054***	−0.051***
	(0.010)	(0.009)	(0.009)
Total Experience	−0.025***	−0.001	−0.019*
	(0.002)	(0.005)	(0.011)
Fraction of Difficult Experience	−0.163***	−0.349***	−0.122***
	(0.012)	(0.03.8)	(0.046)
Number of observations	137,676	137,676	137,676
Log likelihood	−59,366.45	−57,857.61	−57,545.91

Notes. Categorical patient controls include ESI, CCS, and race. Standard errors clustered by physician are in parentheses.
***$p < 0.01$; *$p < 0.1$.

experience increases, physicians get faster at processing patients. Specifically, we find that the coefficient for the cumulative volume of cases is negative (−0.025, $p < 0.01$). We also find that the relative experience gained from taking on difficult cases reduces the service time (−0.163, $p < 0.01$). Specification (2) drops the temporal heterogeneity, and we find that the estimate for the effect of ratio of difficult and easy cases continues to be negative and statistically significant (−0.349, $p < 0.01$). Specification (3) is the full model specification, and it includes the physician fixed effects, temporal factors, and patient-level heterogeneity. We find that the results are qualitatively similar: more experience with patient volume speeds up service rates (estimate = −0.019, $p < 0.01$), and the speed-up effect is more pronounced if a greater share of the experiences are generated from difficult cases (estimate = −0.122, $p < 0.01$). Collectively, the results from Table 6 provide support for the hypothesis that performing more difficult cases helps the physician learn faster than performing easy cases.

Table 7 presents the results for the effects of experience from easy and difficult tasks based on per-patient RVU, an alternative measure of performance. Specifically, we examine the effect of difficult relative to easy experiences on the number of RVUs per future patient. Our hypothesis is that with difficult experience, physicians become more productive, as indicated by the number of RVUs produced per patient. Our results are consistent with the findings for service times. The coefficient estimates across all of the model specifications, which include differing levels of physician and temporal heterogeneity, show that a greater experience in treating difficult cases relative to more easy cases leads physicians to be more productive, as indicated by the amount of work performed per patient. In particular, from our full model specification (specification (3)), we find that the coefficient for total experiences, which has been standardized, is 0.367 ($p < 0.05$). Moreover, we find that the experiences gained from difficult cases helps to increase the productivity on future cases; a greater fraction of experience with difficult cases appears to increase the number of RVUs recorded on future patients (coefficient = 1.645, $p < 0.05$). Collectively, we find that the results from Table 7 are consistent with the results from Table 6, that experiences gained from harder cases lead to greater long-term productivity compared with the experiences gained from easy cases.

In sum, our results show that easy patients are more likely to be selected as workload increases, and this directly affects the shift-level throughput volume

Table 7. Learning Effects on RVUs

	(1)	(2)	(3)
Physician fixed effects		X	X
Time fixed effects	X		X
Female	0.214***	0.191***	0.203***
	(0.041)	(0.035)	(0.035)
Age	–0.007***	–0.009***	–0.007***
	(0.001)	(0.001)	(0.001)
SystemLoad	0.020***	–0.001	0.013**
	(0.005)	(0.005)	(0.006)
PhyLoad	–0.232***	–0.149***	–0.176***
	(0.007)	(0.009)	(0.009)
PhyCasemix	0.571**	–0.477**	0.424*
	(0.230)	(0.218)	(0.222)
WaitLoad	0.014*	–0.007	0.010
	(0.007)	(0.007)	(0.008)
WaitCasemix	–0.385**	0.061	–0.242
	(0.193)	(0.184)	(0.186)
Total Experience	0.060	–0.016	0.367**
	(0.038)	(0.102)	(0.174)
Fraction of Difficult Experience	5.098***	4.084***	1.645**
	(0.339)	(0.604)	(0.809)
Number of observations	118,428	118,428	118,428
Log likelihood	–398,470.83	–397,840.31	–397,546.24

Note. Standard errors clustered by physician are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

(short-term effect). However, TCP leads to fewer RVUs generated per shift. In other words, TCP may be misleading physicians into believing that it improves short-term productivity. In the long run, continued selection of easy cases can limit the physician's ability to learn from difficult tasks, thereby lowering overall baseline productivity. In other words, TCP amounts to "kicking the can down the road" at the detriment of long-term productivity.

4. Study 2: Identifying the Mechanism

Our second hypothesis suggested that individuals focus on easier tasks under increased load because their previous work resulted in (a) lower sense of progress, (b) greater sense of fatigue, and (c) greater sense of stress. Though our field data show support for the fatigue hypothesis, we are not able to test all three explanations. By moving to the laboratory we can explore the mechanisms by asking participants to indicate how they feel about the work they completed before they had an opportunity to choose to work on a different set of tasks. We reasoned that, for example, those who had experienced a higher workload in the first part of the study would feel as if they have not made progress but also were more fatigued and more stressed and thus would choose easier tasks either to experience more sense of progress or to lower their level of fatigue or stress.

Moving to the laboratory also allows us to test the robustness of the effects found in the field data and the generalizability of the TCP. In the laboratory, we did not ask participants to behave similar to physicians dealing with different workloads and choosing patients. Instead, we created a situation in which participants worked on a task that required attention and effort, as well as accuracy. Given that working on multiple patients causes a mental workload for physicians, we varied the mental workload participants experienced in our study.

4.1. Participants

Three-hundred sixty-five adults recruited from Amazon MTurk (59% male) participated in the study in exchange for a $3 payment.[5] We calculated our sample size based on an estimate of an effect size $d = 0.3$, requiring a sample size of approximately 350 participants for a study powered at 80%. We recruited 370 people in the hope of reaching the sample we wanted. We randomly assigned participants to a high-workload or a low-workload condition.

4.1.1. Manipulation of Workload. In this study, as their first task, participants were asked to copy and type up a page of text from a book based on an image of it that they saw on the screen (sideways, to make the task more challenging). Participants worked under time pressure and only had three minutes to type on the screen, in the appropriate window, as much of the text as they could. They were told that accuracy is important, so they were asked to "please try to make as fewer errors as possible as you work on this task."

We manipulated participants' workload by asking half of them (high-workload condition) to listen to a song while copying the text and count how many times five words got mentioned in the song. This manipulation of workload has been previously used in the literature to vary mental workload (Recarte and Nunes 2003). The other half (low-workload condition) did not listen to a song while typing. Thus, those in the high-workload conditions had more work at the same time (typing and listening) than those in the low-workload condition. With this request, an individual is taxed, and one feels busier because it asks people to juggle different tasks. This is similar to our field context because in the case of the doctor, the "various" patients may be treated sequentially, but the thought of them happens "simultaneously" when the pickup decision is made. In other words, in this study, workload is again at the individual level. Although system workload may have broader implications, that is not our focus here. Individuals experience workload (with a given number of patients in Study 1 or with our manipulation in Study 2), and then they have a subsequent choice to make about a pickup decision.

4.2. Procedure

Participants were told that the study consisted of different tasks and short questionnaires. They first completed the first task described above. After completing the typing task, participants were asked to indicate their reactions to it on a 7-point Likert scale (1 = not at all, 7 = very much so). We assessed three measures: perceived sense of progress (three items), perceived fatigue (three items), and perceived stress (three items). The items used to measure sense of progress were as follows: I am feeling (1) as if I am making good progress on the tasks at hand, (2) as if I am progressing well on the tasks at hand, and (3) as if I am completing the work at a good pace. We averaged participants' answers across these three items to create a measure of perceived sense of progress ($\alpha = 0.97$). The items used to measure fatigue were (1) I feel mentally tired, (2) I feel fatigued, and (3) I feel exhausted. We averaged participants' answers across these three items to create a measure of perceived fatigue index ($\alpha = 0.96$). We measured perceived stress with the following items: (1) I feel stressed, (2) I feel anxious, and (3) I feel under pressure; we again averaged answers into a stress index ($\alpha = 0.94$).

Next, participants moved on to the task called "Letter Processing and Word Creation." They were told that the task is focused on demonstrating how people process letters to form words. The instructions informed participants,

> We used this task in other surveys in the past. There are two versions of it. You can choose below the version you'd rather complete.
>
> Version 1: Most people find this version of the task to be quite challenging and had to work hard at it.
>
> Version 2: Most people find this version of the task to be not that challenging and did not have to work hard at it. Which version of the task do you want to complete?

Participants indicated the version of the task they wanted to complete. We used their choice as our primary dependent measure in the analyses reported below.

Independent of their choice, participants moved on to the *same* task. Participants completed seven rounds of the "Letter Processing and Word Creation" task. In each round, they received a set of letters (e.g., "S", "T" "O", "R", C" and "A") and were asked to create as many common English words (excluding proper nouns, such as names of people and places) from these letters as possible and write them in the text box provided. They were asked to try to spend at least a minute or two on each of them. The instructions informed them that, at the end, they would receive a performance score based on the number of words they generated.

Next, participants were asked to think back to the first task they had completed (copying the text) and indicate the extent to which they felt each of three states while completing the task, using a 7-point Likert scale (ranging from 1 = not at all to 7 = very much so): (1) I felt busy, (2) I felt occupied, and (3) I felt I had a high level of workload in my hands. We used these three questions as a manipulation check ($\alpha = 0.89$).

Finally, participants answered a few demographic questions.

4.3. Results

Table 8 shows the descriptive statistics of the variables captured in Study 2 by condition.

4.3.1. Manipulation Check. As expected, participants reported feeling that they had a higher workload in the high-workload than in the low-workload condition ($t(363) = 2.96$, $p = 0.003$). They also transcribed less text in the high-workload than in the low-workload condition ($t(363) = -2.77$, $p = 0.006$). These results suggest our workload manipulation was effective.

4.3.2. Perceived Progress, Fatigue, and Stress. Participants in the high-workload condition reported perceiving a lower sense of progress compared with participants in the low-workload condition ($t(363) = -2.71$, $p = 0.007$), and they felt more fatigued and more

Table 8. Descriptive Statistics of the Variables Captured in Study 2 by Condition

	High-workload condition	Low-workload condition
Perceived workload (man check)	5.59 (1.31)	5.15 (1.50)
Number of words copied	70.02 (37.82)	81.23 (39.47)
Perceived sense of progress	3.93 (1.78)	4.43 (1.70)
Feeling fatigued	3.69 (1.85)	3.00 (1.76)
Feeling stressed	3.64 (1.89)	2.90 (1.69)
Choice of easy task (%)	76.0	63.7
Performance score on task 2	68.14 (44.10)	69.61 (42.05)

stressed ($t(363) = 3.63$, $p < 0.001$ and $t(363) = 3.97$, $p < 0.001$, respectively).

4.3.3. Choosing the Easy Task. As predicted in Hypothesis 1, a higher percentage of participants chose the easy task in the high-workload condition (76.0%, 133/175) than in the low-workload condition (63.7%, 121/190) ($\chi^2(1, N = 365) = 6.53$, $p = 0.011$).

4.3.4. Performance on the Second Task. Independent of the choice made, participants all completed the same seven rounds of the "Letter Processing and Word Creation" task. Thus, we did not expect to find any differences in performance on this task between the high-workload condition and the low-workload condition. This was, in fact, the case ($t(230) = 0.33$, $p = 0.74$).

4.3.5. Mediation Analyses. Our second hypothesis predicted that individual's task selection would be driven by feelings of progress (Hypothesis 2A), fatigue (Hypothesis 2B), and stress (Hypothesis 2C). We found support for progress and fatigue, although stress did not serve as a mediator.

We first ran simple mediation models using the bootstrapping approach outlined by Preacher and Hayes (2004). We estimated the direct and indirect effects of the high-workload condition via each of the proposed mediators (feelings of progress, fatigue, and stress) on our dependent variable: the choice to complete a difficult task as the second task.

On the basis of a bootstrapping (with 10,000 iterations) analysis, our manipulation of high workload had a significant effect on sense of progress (B = −0.49, se = 0.18, $p = 0.007$), which, in turn, significantly affected the choice of easy task (B = −0.22, se = 0.07, $p = 0.002$). The 95% bias-corrected confidence interval for the size of the indirect effect excluded zero for sense of progress [0.024, 0.261], suggesting significant mediation. For fatigue, similarly, the workload condition had a significant effect on fatigue (B = 0.69, se = 0.19, $p < 0.001$), which, in turn, significantly affected the choice of easy task (B = 0.21, se = 0.07, $p = 0.003$). The 95% bias-corrected confidence interval for the size of the indirect effect excluded zero for fatigue too [0.045, 0.306], suggesting significant mediation. Finally, our manipulation of workload significantly affected stress (B = 0.74, se = 0.19, $p < 0.001$), which, in turn, affected the choice of easy task (B = 0.18, se = 0.07, $p = 0.009$). The 95% bias-corrected confidence interval for the size of the indirect effect excluded zero [0.035, 0.296], once again suggesting significant mediation.

We then turn to assess and compare indirect effects in a multiple mediator model (i.e., simultaneous mediation by multiple variables) using the bootstrapping approach outlined by Preacher and Hayes (2008). In a multiple-mediator model, the mediators' unique abilities to mediate, above and beyond any other mediators or covariates in the model, are tested. Results with 10,000 bootstrap samples provided significant support for a model in which feelings of progress and fatigue serve as mediators for the relationship between the high-workload condition and the choice to complete an easy task as the second task. The 90% bias-corrected confidence interval for the size of the indirect effect excluded zero for both mediators ([0.018, 0.212] and [0.013, 0.267], respectively), suggesting marginally significant mediation by the two of them. Instead, the 90% bias-corrected confidence interval for the size of the indirect effect included zero for stress [−0.126, 0.123], suggesting that stress does not mediate the effect of workload on the choice of an easy task. We note that, for this mediation analysis, we moved to a 90% bias-corrected confidence interval, as the effect of fatigue on the choice of an easy task, when also including sense of progress and stress as potential mediators, was significant at the 10% level (B = 0.16, se = 0.09, $p = 0.086$). The effect of sense of progress was also significant (B = −0.18, se = 0.08, $p = 0.02$), whereas the effect of stress was not (B = 0.002, se = 0.10, $p = 0.98$).

In an analysis with only sense of progress and fatigue as potential multiple mediators, we found evidence for mediation as well, as the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero for both mediators ([0.011, 0.231] and

[0.018, 0.267], respectively), suggesting significant mediation by the two of them.

4.4. Discussion

The results of Study 2 not only conceptually replicate our first hypothesis but also show that, as predicted in our second hypothesis, a sense of a lack of progress and perceived fatigue after the first task explain why high workload leads to choosing easy versus difficult tasks.

5. Discussion and Conclusion

In this paper we study how individuals manage tasks under varying workload conditions and then investigate the short-term and long-term productivity implications. In so doing we make several contributions to the literature. First, using both the field and the laboratory, we provide evidence of a task completion preference that as individuals experience higher levels of load, they select easier tasks. This finding is important because it provides a new explanation through which workload may impact performance. Although prior literature has shown that higher levels of load may lead to improved performance (KC and Terwiesch 2009), at least to a point, here we show another reason why high load may impact performance—task selection.

This finding is related to prior work on discretionary effort allocation, such as the speed-quality trade-off (Hopp et al. 2007, Anand et al. 2011, Tan and Netessine 2014). However, here, instead of trading off speed and quality within the same type of task, individuals trade off the task that they complete. This creates a number of opportunities for future work. Empirical work should seek to study not only task selection in more detail but also how task selection and the speed-quality trade-off may interact. In addition, studying individual differences in the likelihood to engage in the behavior could yield interesting insights. Analytical work should investigate how the task completion preference may lead to negative outcomes for certain types of work (e.g., harder tasks) while advantaging other types of work (e.g., easier tasks).

Our second contribution comes from investigating whether task completion preference is a valid variance reduction strategy. As expected, we find that TCP is a strategy that helps to manage the variability of service times during a shift. The question then becomes how it affects individual productivity, and therein lies our next finding.

Our third contribution investigates the impact of TCP on throughput volume. Interestingly, it appears at first that TCP improves shift-level throughput volume. We find this result when we simply measure work as the number of patients cared for. However, when we more precisely measure work, using a complexity adjustment (RVUs), we actually find that TCP is related to lower throughput volume. Future work should explore this in more detail. In some contexts TCP may result in pure short-term benefits. An interesting implication of our work is that individuals may be "tricked" into thinking that their performance is improved when they select easier patients because they are discharging more patients than their difficult patient–serving peers. Additional research can consider such elements where workers may inadvertently decrease their own performance through their work execution strategies.

Our fourth contribution investigates the longer-term impact of the task completion preference. Although prior work suggests that improved performance may eventually hinder performance as a result of overwork (KC and Terwiesch 2009, Staats and Gino 2012, Kuntz et al. 2015), here we show a negative learning effect arising from completing easier tasks. The finding is similar to the general idea of exploration and exploitation (March 1991). By selecting the easier task (exploitation), an individual gets work done quicker—and likely feels good doing it. However, by choosing the harder task (exploration), one creates an opportunity to learn. By completing more difficult tasks, the individual identifies the potential learning curves nested within learning curves (Zangwill and Kantor 1998), thus improving performance. Although always selecting the harder task may be suboptimal, if one continually chooses the easier exploitation path, then longer-term performance suffers. Future work should further consider the balance between easy and hard tasks as a key dimension for learning.

Finally, in our work we are, to our knowledge, the first to show the mechanisms through which TCP affects performance—the positive feelings that accrue as work is finished and fatigue. There continues to be an opportunity to understand the mechanisms through which operational performance can be impacted. The benefit of not only documenting a relationship, as we do in this case with task selection and operational performance, but also showing the mechanisms through which it occurs allows us to design systems around a positive mechanism (or avoid a negative one). Here, our findings suggest that understanding ways to get the "completion high" or avoid fatigue (e.g., by breaking tasks into smaller pieces or providing breaks) would be valuable to theory and practice.

5.1. Limitations

As with any study, our study has its own limitations. First, although our field site provides several years of data and many thousands of observations, it is only one hospital. Future work should seek to replicate and extend our findings to other settings and more organizational units. Second, our study examines picking up a hard or easy patient, conditional on picking up a patient at all. Further work should identify areas where the decision to pick up, at all, could be studied. Third, in the field setting of the hospital, we are only able to test the fatigue mechanism. The laboratory provides an opportunity to gain more precise control for testing purposes. However, future work should seek to study all of our identified mechanisms, as well as others, in the field. Fourth, although we consider multiple, precise measures of operational performance, we are not able to evaluate quality with our data. Future studies should consider TCP and its quality implications. Fifth, we look at the time taken to discharge individual patients, which we believe represents an intuitive and effective overall measure of productivity. However, we do not observe detailed processing of the various steps involved in care. In particular, microtask times (including face-to-face interaction, diagnosis, and ordering of tests) are not examined. Future research, based on more granular microlevel timestamp data, could explore further microfoundations of productivity. Finally, we rely on laboratory subjects to identify our mechanism. This is an established and valid practice when studying human decision making because we are interested in a general behavior of engaging in the task completion preference. Still, future work could consider creative ways to engage real participants—for example, as Tucker (2016) had done when she ran experiments on nurses at a convention. Fundamentally, our study relies on triangulation for its validity, but the more work—both on the findings and on the mechanisms—that can be done in the field with practitioners, the more benefits for theory and practice.

5.2. Managerial Implications

Our findings offer important implications for not only theory but also for practice. Prior work highlights the value that accrues when workers have discretion with which to manage their tasks (Bowman 1963, van Donselaar et al. 2010, Ibanez et al. 2018). However, given the use of discretion in constructing an operating system, it is also important to understand how individuals choose their work to complete. In the opening quote of this paper, we note how Frederick Taylor prioritized the idea of building a system to help people develop. In identifying the task completion preference, we find a way that people may work against their long-term self-interest.

In our study we are able to quantify the performance effects as well as the effects of TCP. How then should a manager use our findings? The first step is to appreciate that workload may change not only the speed at which individuals work but also the tasks that they choose to work on. As a result, managers may wish to think about different instructions on how to select tasks when things are busy, or alternatively, they can educate workers about task completion preference as a means to address it. In addition, given the performance implications, managers should educate workers about the performance costs of avoiding hard tasks. This may help workers to continue to develop, as Taylor suggests. Finally, managers can use the finding that completing tasks leads to feelings of progress to structure work effectively. As noted, this could mean taking harder tasks and breaking them into components, each with a clearly defined indicator of completion, so that completion occurs more often. An alternative could be to consider schedule rotation whereby individuals are rotated between easy and hard tasks over time.[6] Because scientific management the structure of work has been a primary focus of operations. That is no less true today—if anything, it has become even more a focus because managers have more degrees of freedom to structure work—sending different tasks to different people, breaking up the work into different pieces, etc. That means managers and academics need to be experimenting on how work should be structured to find better ways.

5.3. Conclusion

The word *operations* is derived from a Latin word for work (Terwiesch 2017). Our field studies many things, but work is central to what we do. Work is done by individuals, and so understanding people-centric operations is necessary for progress in theory and practice. In this paper we draw on research on workload, individual discretion, and decision making to study how workload and task selection interact. We theorize and test in a hospital that, under increased workload, individuals choose to complete easier tasks. We call this behavior task completion preference. We then investigate the performance implications of task completion preference. In addition, in the laboratory we conceptually replicate the task selection effect and show two mechanisms—sense of progress and fatigue—that explain why it occurs. With these findings we identify an additional reason for the workload-speedup effect found in the literature. By better structuring work, it is possible to aid individual development and organizational performance.

Endnotes

[1] Although the work of Taylor (1911) was central to scientific management, his treatment of others different from him is not acceptable by modern standards. We note that now operating systems should seek to develop "first-class humans."

[2] Ibanez et al. (2018) examine whether a task is on average an individual's shortest task. There is heterogeneity in the shortest task

across radiologists, and they do not investigate whether a task is easier or harder than another task.

[3] See https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp, accessed May 8, 2019.

[4] See https://www.nhpf.org/library/the-basics/Basics_RVUs_01-12-15.pdf for more details (last accessed May 8, 2019).

[5] A similar conceptual replication was run in a university laboratory, which included a sense of progress but not fatigue and stress as mediators.

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

References

Amabile T, Kramer S (2011) *The Progress Principle: Using Small Wins to Ignite Joy, Engagement, and Creativity at Work* (Harvard Business Review Press, Boston).

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.

Arlotto A, Chick SE, Gans N (2014) Optimal hiring and retention policies for heterogeneous workers who learn. *Management Sci.* 60(1):110–129.

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

Batt RJ, Terwiesch C (2017) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Sci.* 63(11):3531–3551.

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.

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

Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and students as newsvendors. *Management Sci.* 58(12):2225–2233.

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

Buell RW, Kim T, Tsay C-J (2017) Creating reciprocal value through operational transparency. *Management Sci.* 63(6):1673–1695.

Cachon G, Terwiesch C (2009) *Matching Supply with Demand* (McGraw-Hill Irwin, Boston).

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

Chan DC Jr (2018) The efficiency of slacking off: Evidence from the emergency department. *Econometrica* 86(3):997–1030.

Crabill TB (1972) Optimal control of a service facility with variable exponential service times and constant arrival rate. *Management Sci.* 18(9):560–566.

Csikszentmihalyi M (1996) *Creativity: Flow and the Psychology of Discovery and Invention* (HarperCollins Publishers, New York).

Dai H, Milkman KL, Hofmann DA, Staats BR (2015) The impact of time at work and time off from work on rule compliance: The case of hand hygiene in healthcare. *J. Appl. Psych.* 100(3):846–862.

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.

Debo LG, Toktay LB, Van Wassenhove LN (2008) Queuing for expert services. *Management Sci.* 54(8):1497–1512.

Deo S, Jain A (2019) Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production Oper. Management* 28(5):1061–1081.

Freeman M, Savva N, Scholtes S (2017) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Sci.* 63(10):3147–3167.

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

Gerchak Y, Gupta D, Henig M (1996) Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Sci.* 42(3):321–334.

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

Grant AM, Fried Y, Parker SK, Frese M (2010) Putting job design in context: Introduction to the special issue. *J. Organ. Behav.* 31(2–3):145–157.

Green LV, Savin S, Savva N (2013) "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* 59(10):2237–2256.

Green LV, Savin S, Wang B (2006) Managing patient service in a diagnostic medical facility. *Oper. Res.* 54(1):11–25.

Hackman JR, Oldham GR (1976) Motivation through the design of work: Test of a theory. *Organ. Behav. Human Performance* 16(2):250–279.

Hayes RH (2002) Wick Skinner: A life sailing against the wind. *Production Oper. Management* 11(1):1–8.

Heath C, Larrick RP, Wu G (1999) Goals as reference points. *Cognitive Psych.* 38(1):79–109.

Heilizer F (1977) A review of theory and research on the assumptions of Miller's response competition (conflict) models: Response gradients. *J. General Psych.* 97(1):17–71.

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

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

Ibanez MR, Clark JR, Huckman RS, Staats BR (2018) Discretionary task ordering: Queue management in radiological services. *Management Sci.* 64(9):4389–4407.

KC D (2013) Does multi-tasking improve productivity and quality? Evidence from the emergency department. *Manufacturing Serv. Oper. Management* 16(2):168–183.

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.

KC D, Scholtes S, Terwiesch C (2020) Empirical research in healthcare operations: past research, present understanding, and future opportunities. *Manufacturing Service Oper. Management* 22(1):73–83.

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.

Kivetz R, Urminsky O, Zheng Y (2006) The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *J. Marketing Res.* 43(1):39–58.

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

Kuntz L, Mennicken R, Scholtes S (2015) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Sci.* 61(4):754–771.

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

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

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.

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

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.

Preacher KJ, Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instruments Comput.* 36(4):717–731.

Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40:879–891.

Recarte MA, Nunes LM (2003) Mental workload while driving: Effects on visual search, discrimination, and decision making. *J. Experiment. Psych. Appl.* 9(2):119–137.

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, 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.

Smiddy H, Naum L (1954) Evolution of a "science of managing" in America. *Management Sci.* 1(1):1–31.

Song H, Tucker A, Morrell K (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Sci.* 61(12):3032–3053.

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

Staats BR, KC DS, Gino F (2018) Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Sci.* 64(2):804–824.

Staw BM, Sandelands LE, Dutton JE (1981) Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis. *Admin. Sci. Quart.* 26(4):501–524.

Stidham S Jr, Weber RR (1989) Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Oper. Res.* 37(4):611–625.

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).

Terwiesch C (2017) *MSOM Fellows talks.* Lecture, Manufacturing & Service Operations Management Conference, June 21, Manufacturing and Service Operations Management Society, INFORMS, Catonsville, MD.

Tucker AL (2016) The impact of workaround difficulty on frontline employees' response to operational failures: A laboratory experiment on medication administration. *Management Sci.* 62(4): 1124–1144.

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.

Weick KE (1984) Small wins: Redefining the scale of social problems. *Amer. Psych.* 39(1):40–49.

Zangwill WI, Kantor PB (1998) Toward a theory of continuous improvement and the learning curve. *Management Sci.* 44(7): 910–920.

Zipkin PH (2000) *Foundations of Inventory Management* (McGraw-Hill, New York).