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

Diwas S. KC

Maryam Kouchaki

Bradley R. Staats

Francesca Gino

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

Diwas S. KC
Emory University

Bradley R. Staats
University of North Carolina at Chapel Hill

Maryam Kouchaki
Northwestern University

Francesca Gino
Harvard Business School

Working Paper 17-112

Task Selection and Workload:

A Focus on Completing Easy Tasks Hurts Long-Term Performance

Diwas S. KC
Emory University
1300 Clifton Road
Atlanta, GA 30322
Tel: 404.727.1424
diwas.kc@emory.edu

Bradley R. Staats
University of North Carolina at Chapel Hill
Campus Box 3490, McColl Building
Chapel Hill, NC 27599-3490
Tel: 919.962.7343
bstaats@unc.edu

Maryam Kouchaki
Kellogg School of Management
Northwestern University
2001 Sheridan Rd.,
Evanston, IL 60208
Tel: 847.491.4379
m-kouchaki@kellogg.northwestern.edu

Francesca Gino
Harvard Business School
Harvard University, Baker Library
Boston, MA 02163
Tel: 617.495.0875
fgino@hbs.edu

June 25, 2017

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

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 increase their service time, up to a point, in order 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 in order to manage their load. We label this behavior Task Completion Bias (TCB). Using two years of data from a hospital emergency department we find support for TCB and also show that it improves short-term productivity. However, although it improves performance in the short-term we find that an overreliance on this task selection strategy hurts performance – as measured both by speed and revenue – in the long run. We then turn to the lab to replicate conceptually the task selection effect and show that it occurs due to the positive feelings individuals get from task completion. These findings provide an alternative mechanism for the workload-speedup effect from the literature. We also discuss implications for both research and the practice of operations in building systems to help people succeed in both the short and long run.

Key Words: 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."

– Frederick Taylor (1911: 2)

1. Introduction

Since its roots in the Scientific Management movement that arose in response to the increasing operational complexity of the 2nd Industrial Revolution the field of operations management has sought to improve system design in order to better match supply to demand (Smiddy and Naum 1954). Throughout the 20th Century, that meant improvements in the structure of how work was accomplished, such as better work scheduling, improved inventory models, or enhanced routing of calls to agents (e.g., Zipkin 2000; Gans, Koole and Mandelbaum 2003; Pinedo 2012). With few exceptions, such as the work of Wickham Skinner (Hayes 2002), when people were

considered in these models they were understandably treated as fixed entities that needed to be scheduled or otherwise manipulated.

As the quote from Taylor above notes, however, people are different from the inanimate objects – e.g., inventory or machines – that the field of operations often focuses its attention upon. People can develop, or change, as a result of the system in which they operate. Individuals respond to stimuli and alter their behavior – sometimes to improve the system and sometimes to detract from it. The role of individuals within operations grows more central as we see an ongoing increase in service and knowledge operations where the individuals and their ability to learn and adapt often serve as a primary source of competitive advantage. Not surprisingly then a large body of scholars has embraced this focus on People Operations. Findings have highlighted the critical role of factors such as learning (Lapré and Nembhard 2010; KC, Staats and Gino 2013), individuals subverting the system through workarounds (Tucker 2007; 2015), employee interaction with customers (Buell, Kim and Tsay 2016), and the choice of whether to execute a given task, or not (Freeman, Savva and Scholtes 2016; Ibañez et al. 2017).

One of the early findings within the People Operations literature was that the rate at which individuals work is not exogenously determined. Using multiple settings within healthcare KC and Terwiesch (2009) show that service rates are endogenous to load. Multiple papers have built upon this finding: to replicate it in other contexts (Staats and Gino 2012), to show that quality may suffer due to load (Kuntz, Mennicken and Scholtes 2015), to show that workers may burn out due to load (Green, Savva and Savin 2012) and more generally show how load can alter behavior in an operating system (Tan and Netessine 2014; Berry Jaeker and Tucker 2015; Kim et al. 2015). A key assumption in this line of work is that as individuals experience more load, they choose to work faster in the short-term, although this speeding up may negatively impact performance in the long-term.

In this paper we offer a different explanation as to why performance may improve as workload increases – task selection. Recent work in People Operations has highlighted that individual discretion – a person's decision about how to alter her work – may have important operational consequences – sometimes leading to improved performance (van Donselaar et al. 2010; Campbell and Frei 2011; Kim et al. 2015; Phillips, Şimşek and Ryzin 2015) and sometimes leading to worse performance (Ibañez et al. 2017). For example, Ibañez et al. (2017) find that, on average, individuals choose tasks in a queue that have the shortest completion time

and that this changing of the order of tasks leads to slower, overall task completion times. In this paper we build on the operations' work considering the role of discretion, as well as work that suggests that people may exhibit a bias towards completing tasks over choosing the appropriate task (Amabile and Kramer 2011; Amar et al. 2011), in order to examine whether individuals alter their task selection when workload increases. We propose that people will show a task completion bias (TCB) 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, as compared to states of lower workload. We consider whether this affects both short-term and long-term performance.

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 two years of data, over 90,000 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, and so as econometricians see the same information that the doctor saw when she made her patient pickup decision. We can then analyze not only the short-term impact on productivity, but also the longer-term impact.

To further understand the decision making process for individuals, we then turn to the lab. The controlled, research laboratory is an important and helpful methodology to couple with field data as it permits us to not only conceptually replicate our main finding, but also to understand the mechanism through which it occurs. Increasingly we find that People Operations studies are combining these two approaches (e.g., Buell et al. 2016; Staats, KC and Gino 2016).

In both the lab and the field we find that, on average, people show a task completion bias as load increases. In other words, we find that when the level of workload increases, workers systematically select easier tasks over difficult tasks, exhibiting TCB. This is important since we know that how individuals manage and process their workload can have significant implications for the productivity of workers in the modern workplace. Using the lab we are able to examine the mechanism by which task completion impacts performance as we find that completing tasks leads individuals to feel good and that increases short-term performance. In other words, it makes workers feel good to complete the tasks, even if the tasks are easy. However, we find that this positive feeling may be misleading. In the immediate short term, picking easier tasks is associated with a higher rate of productivity. However, when we examine long-term

productivity, workers who exhibit TCB tend to be significantly less productive than workers who do not exhibit this bias. This finding is robust across different measures of productivity. Interestingly, we find that workers who exhibit TCB tend to have lower variability in their overall task completion times. We believe that by picking up easier cases when workload increases, TCB workers prevent the overall task completion times from significantly increasing during periods of high workload. This means that even though TCB hinders long-term productivity, it does offer a temporary relief during high-workload periods, leading to more standardized overall task completion times.

Overall our study makes several contributions to the literature. First, we provide evidence of task completion bias, which offers an alternative explanation for the performance effects of workload. Second, we show that there are system benefits, in the short-term when workers select these easier tasks as they complete work faster. Third, although prior work suggests that improved performance may eventually hinder performance due to overwork (KC and Terwiesch 2009; Staats and Gino 2012; Kuntz et al. 2015) here we show a negative longer-term effect that is consistent with a lack of learning arising from completing easier tasks. Finally, as a part of these models we are the first to show the mechanism through which completion benefits performance – 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 literature in operations that has considered the sizing of capacity and then the scheduling of servers to complete the work (e.g., Crabill 1972; Stidham Jr and Weber 1989; Pinedo 2012). This literature covers a wide variety of different problems – from allocating capacity (Green, Savin and Wang 2006) to sequencing work (Gerchak, Gupta and Henig 1996) to scheduling service appointments (Bassamboo and Randhawa 2015). KC and Terwiesch (2009) contributed 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 seminal paper. For example, Tan and Netessine (2014), drawing on literature that shows a speed-quality tradeoff in discretionary services (Hopp, Iravani and Yuen 2007; Debo, Toktay and Van Wassenhove 2008; Anand, Paç and Veeraraghavan 2011), hypothesize that as workload increases individuals may alter not only the speed of the service that they offer – as seen in KC and Terwiesch (2009), but also the quality of the service. Using a sample of restaurant servers, the authors support their hypothesis finding an inverted U-shaped relationship between workload and meal duration as servers adjust their service quality to maximize overall revenue. Kuntz et al. (2015) add to the literature by focusing attention 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 shows an increase. Also examining hospital data Berry Jaeker and Tucker (2015) find that beyond a certain level of congestion in the system, patient length of stay increases because the patients that are left in the system have high demands. Other papers have considered the role of workload, for example in the intensive care admissions decision (Kim et al. 2015) and in emergency department service times (Batt and Terwiesch 2016). We contribute to this line of literature as we are the first to consider, directly, the role of individual task selection in the workload speedup effect.

Second our work builds on literature that has looked at 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 simply 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 is a topic that has been examined from several different dimensions in the operations literature such as routing work to different servers (Shumsky and Pinker 2003; Saghafian et al. 2014; Freeman et al. 2016), capacity allocation (Kim et al. 2015), making a tradeoff between speed and quality (Hopp et al. 2007; Anand et al. 2011; Powell, Savin and Savva 2012), working in a dedicated vs. a pooled queue (Song, Tucker and Morrell 2015), and determining processing times in the face of different inventory levels (Schultz et al. 1998; Schultz, Juran and Boudreau 1999).

The literature has found that worker's discretion – whether in the examples in the previous sentence or the workload examples in the previous paragraph – has a meaningful impact on operational outcomes. For example, Ibañez et al. (2017) find that radiologists are likely to

6

reorganize their queue of work – 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 they meaningfully slow down their performance. We contribute to this line of research by further unpacking the role of discretion in operating systems. Our work is most similar to Ibañez et al. (2017), but whereas they look at how individuals manage an individual queue, we consider how individuals select tasks from a common queue under different workload conditions. 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 looks at how individuals may deviate suboptimally, from decision models that describe rational agents. Both Bendoly, Donohue and Schultz (2006) and Gino and Pisano (2008) provide reviews of this literature. Biased decision making – decisions that deviate from the rational choice have been shown in numerous operational domains such as forecasting (Kremer, Moritz and Siemsen 2011), contracting (Davis, Katok and Santamaría 2014) and inventory management (Schweitzer and Cachon 2000). Perhaps most related to our work is research by Amar et al. (2011) that shows individuals may pay back smaller, lower interest debts, rather than portions of larger, higher interest debts, in order to close out or 'complete' the smaller debts. In this sense, individuals tend to discount task complexity, while demonstrating a greater importance on the number of tasks that have been completed. We contribute to this work by focusing on attention on operational tasks, and in so doing identifying the task completion bias, but also we explore two possible mechanisms – a positive feeling from completion and a sense of progress that may drive this effect. We now turn our attention to this paper's hypotheses.

2.2 Task Selection and 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, as compared to low levels of workload. There are three theoretical reasons to think that is the case. The first comes from the fact that people seek to make progress in the work that they do. This has been captured in the goal gradient hypothesis which states that individual motivation increases as goal completion draws near (Heilizer 1977). Kivetz, Urminsky and Zheng (2006) show the effect with consumers finding 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 customers given a card requiring 10 stamps

(and having zero punched). Deo and Jain (2015) and Chan Jr. (2015) both find evidence of a goal gradient effect in the healthcare setting as providers work faster near the end of their shift.

Here we are suggesting that the goal gradient effect may impact individual's task selection. When choosing between multiple tasks, individuals may choose the easiest in order to make progress towards their overall goal – finishing the work or at least decreasing the amount of work still to be done. Ibañez et al. (2017) show that when radiologists deviate, or change the order of tasks in their queue, one of the reasons they deviate is to switch to the shortest task.[1] In our study we are interested in how task selection may change due to workload. On average a focus on completing tasks may exist due to the goal gradient effect and the desire to make progress. We hypothesize though that this effect is amplified with increasing workload. It is likely that the desire for progress may increase as workload increases. Prior work notes that when an overall goal is challenging – such as managing a large amount of work – then individuals may look to break the work into smaller pieces that can be addressed more easily (Newell and Simon 1972). In addition, individuals are likely to be spread more thinly during periods of high workload; as a result, workers may not have the additional cognitive resources needed to take on additional complex tasks. It is likely then that with higher levels of workload, individuals will be more likely to look for ways to make progress in their work and so may choose easier tasks that can be completed more quickly.

The second reason that individuals may focus on task completion is that completing work makes people feel good (Amabile and Kramer 2011). When we complete tasks we get a feeling of positive emotion. This positive feeling is something that we anticipate and it may motivate us to both work harder and seek the experience again (Weick 1984).

And so, again, we would expect to see a completion focus, on average. But why might a focus on completion increase when workload increases? When the amount of work to be completed increases, then individuals are likely to feel more anxiety or stress. Prior work shows that anxiety changes how individuals view the situation that they are in, namely they view the situation as more of a threat (Staw, Sandelands and Dutton 1981). With this change in perspective individuals are likely to restrict their information processing and attend to fewer signals and use simpler rules as they decide which task to select next in order to most effectively

---

[1] Note that Ibañez et al. (2017) look at whether a task is on average the shortest task for an individual radiologist. There is heterogeneity in the task that is the shortest across radiologists and so they do not explicitly look, or try to look at whether a task is easier or harder than another task.

process the work. As a result, they may select easier tasks to make progress, but also they may decide to choose these tasks as the positive feelings from task completion may help 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 in order to minimize average wait time one should complete the shortest task first (Cachon and Terwiesch 2009). Although an individual might choose to follow this rule at any point, as workload increases then wait times may grow more salient and so individuals may default to this rule. As a result of making progress, positive feelings from task completion, and a focus on wait times we hypothesize:

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

2.3 Mechanism of Task Completion Bias

As our second hypothesis we turn to the potential mechanism that may drive the effect. Above we highlighted two possible mechanisms: (1) sense of progress; (2) good feelings. Prior literature talks about both and does not specify which is more likely to drive the effect. It is plausible that both could be at play or alternatively one could dominate. Therefore, we offer each as independent hypotheses:

HYPOTHESIS 2A: *People are more likely to select easy tasks than harder tasks when they are confronted with 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 they are confronted with a high versus a low workload because it makes them feel good.*

2.4 Performance Consequences of Task Completion Bias

We now turn to the performance consequences of selecting easier tasks. Whether it is done for psychological or operational reasons, the expectation is that choosing the easier task will result in faster completion of the task, helping to manage the workload. This general decision making process is consistent with Bowman (1963) and the idea of the management coefficients theory. In this perspective, individuals' actual choices are likely to show improvements over basic models because individuals have unique information to better inform

their decisions. For example, van Donselaar et al. (2010) find that supermarket store managers deviate from the company's inventory models as they recognize problems with the system and in so doing they improve overall performance. However, Ibañez et al. (2017) use a similar motivation and find that deviating from the standard order of tasks has negative performance implications.

The question then becomes what would one expect in this context? Here the comparison is that when an individual chooses an easy task she is *not* executing a harder task. Part of the motivation for selecting the easy task is that it is likely to be quicker. In Ibañez et al. (2017) the likely reason that changing tasks led to slower performance is that changing required a time for searching and not changing did not require a search cost. In settings where selecting an easy versus hard task both require a similar search cost then the former view – easy tasks likely take less time should dominate. As a result we hypothesize:

HYPOTHESIS 3:   *Selecting easy tasks over hard tasks results in short-term productivity improvements.*

Although selecting an easy task over a hard task may help to manage short-term productivity in the face of increased workload, it may lead to longer term consequences. The topic of experience and its role on performance and choices has received a great deal of attention in the operations literature (Lapré and Nembhard 2010; Bolton, Ockenfels and Thonemann 2012; KC and Staats 2012; Arlotto, Chick and Gans 2014). Traditionally experience is seen as beneficial – as individuals accrue experience then they learn and improve – although experience can lead to suboptimal choices in some cases (Staats et al. 2016). Recent work on experience highlights that not all experience is created equal and that some types of experience may be more beneficial for performance and learning (Huckman, Staats and Upton 2009; Narayanan, Balasubramanian and Swaminathan 2009; Staats and Gino 2012).

Here our question is whether the consistent selection of easy tasks might lead to worse, or better, longer term performance? Learning from experience is premised on a learning curve – with repeated experience an individual improves at a task (Lapré and Nembhard 2010). Some of that learning comes from learning a routine, but much of it comes from learning the intricacies of the task. The challenge with this is that by consistently focusing on the easy tasks an individual may miss out on learning opportunities – not challenging herself to learn more deeply on the 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, Larrick and Wu 1999; Amar et al. 2011). And so, here we hypothesize that as individuals focus on completion and easy tasks then even if there are short-term gains, their long-term productivity is likely to suffer:

HYPOTHESIS 4:   *In the long term, selecting easy tasks over hard tasks results in significant declines in productivity.*

We now turn our attention to testing our hypotheses in the field and in the lab. We first go to the field in order to examine whether individuals exhibit a task completion bias by choosing easy tasks over hard tasks (H1) and then the short-term (H3) and long-term (H4) productivity implications. We then go into the lab to replicate conceptually the task completion bias finding (H1) and investigate the potential mechanism (H2a and H2b).

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

In this study we investigate whether task completion bias is exhibited by professional knowledge workers in the workplace. We specifically look at the task selection behavior of physicians in the emergency department (ED) of a busy hospital, and examine any resulting effects of short-term and long-term productivity.

The ED provides an ideal field setting in which to study our research questions for several reasons. First, the arrival of patients to the ED is an inherently random occurrence. The random arrival of patients to the ED means that 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 means that we observe appreciable variation in the physician workload, which we can use to identify the task selection effect. Finally, the time taken by the physician in discharging the patient is clearly defined. This allows us to accurately estimate the productivity of individual physicians.

In the context of the ED, easy tasks correspond to the treatment of patients with lower levels of acuity. We postulate that under increased levels of workload, physicians are more likely to pick up low-acuity patients, rather than high-acuity patients, consistent with a task completion bias. We also postulate that in the short term, these low-severity patients are

associated with shorter lengths of stay, which correspond to higher productivity. In the long term, we hypothesize that TCB is associated with lower productivity.

3.1 Setting

The field study context is the emergency department (ED) of an East Coast metropolitan hospital that treats a large 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 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 given 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 for providing updates to the clinical record upon treatment. Physicians continuously monitor the queue to pick up new patients when not attending to their 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 an evaluation of the patient's medical record, physical examination of the patient, gathering of relevant medical information, ordering tests and procedures, and a final disposition decision involving a discharge from the ED. At any given 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 variation in physician workload is used to identify the physician's choice of whether to pick up an easy or a difficult patient.

3.2 Data

We assembled our data from the emergency department for twenty-four months in fiscal years 2006-2007 involving over 90,000 distinct patient encounters. Our data include patient-level visit information, as well as physician level and ED-level factors. Most importantly, 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 individual patient, as well as the productivity of each individual physician.

Our data set includes various measures of micro-level productivity. In particular, we observe the time a patient is picked up by a physician and when she is discharged. The time that

a patient is under the care of a given physician (physician assignment period) 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, which we denote the service time (SVCTIME) is our primary outcome measure of productivity.

The micro-level operational data allow us to construct a time series for the number of patients seen by a given physician at a specific point in time. 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 value of workload remains constant until either a pick-up or discharge event occurs.

The arrival and discharge times also allow us to construct the processing load in the ED at specific points in time. The system load (or occupancy) in the ED at time $t + dt$ is given by the relationship:

$$SystemLoad(t+dt) = SystemLoad(t) + Arrivals_{t,t+dt} - Discharges_{t,t+dt},$$

where $Arrivals_{t,t+dt}$ and $Discharges_{t,t+dt}$ denote the number of arrivals and departures between t and $t+dt$ respectively. This simple formula, in conjunction with the micro-level time stamp data, allows us to construct minute-by-minute $SystemLoad$ for the ED during the entire period of study. Our data also include system-level variable the number of emergency room physicians who are presently working in the ED at time t. Table 1 provides summary statistics for the data set. We find that the average census in the ED is 15.51. However, there is significant variability as indicated by the high standard deviation. The average physician workload is 4.3 patients, and the standard deviation is 2.6 patients.

In addition to the timestamps associated with each patient visit, we also observe patient-level variables, including the patient's acuity level, which is coded as a binary variable to denote whether the patient is of high acuity level or not. In addition, we observe various sources of heterogeneity at the patient level, including age, gender, race, payment status, means of arrival, as well as temporal considerations including time, day of week, and month of year. We find that 18% of the patients are low acuity patients. The average age of patients is 38, but there is significant variation as indicated by the high standard deviation. Approximately 18% of patients are on Medicaid.

Our primary objective is to examine the pickup behavior of the physician as a function of the extent 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 pickup by a physician at any given point in time. At different points in time, individual physicians may have different levels of workload, and this could impact the physician's decision to pick up an easy patient or a more acute patient.

Our second objective is to examine the effect of workload on the physician's productivity, using the individual patient's service time as the outcome measure, which is estimated from the pickup and the discharge timestamps for each patient visit. In addition, since we observe the acuity level of the patient as well as the physician's workload, we can examine the extent to which the processing of low-severity patients can increase physician productivity during periods of increased workloads.

An alternative measure of the productivity associated with the discharge of a given patient is the relative value unit (or RVU) designation for that patient's visit. Specifically, the RVU is a measure of the physician's work associated with the care of a given patient, and takes into account the physician's time, physical and mental effort, as well as clinical and technical judgment and skill. RVU's are tied to the primary diagnosis code for the patient, and are currently used by payers, including Medicare, to reimburse physicians for services rendered. As such, RVU's allow us to examine the implications of TCB on hospital revenues. The average service time is 3.2 hours, while the average RVU is 5.15 units.

By combining the system-level, and physician level factors, in conjunction with the clinical considerations at the patient level, we have at our disposal 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 and the resulting short-term and long-term productivity of ED physicians.

3.3.1 Physician Task Selection

This subsection presents our empirical model of physician choice. Our modeling approach uses individual physician choice of patient to pick up (i.e. revealed preference data) as well as a

number of physician, patient, and emergency department level data 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 amongst alternatives using a logit model. In particular, we are interested in the effects of workload on the physician's choice. Under the logit specification, the probability that physician i chooses patient j at time t is given by:

$$P_i(EASY_j) = \frac{e^{\alpha_i + X_j\beta + Z_t\gamma + \delta SystemLoad_t + \theta Workload_{it}}}{\sum_{ijt} e^{\alpha_i + X_j\beta + Z_t\gamma + \delta SystemLoad_t + \theta Workload_{it}}}$$

$\mathbf{X_j}$ is 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 are determined after the patient is fully evaluated by the physician, and are not available to the physician prior to pick up. $\mathbf{Z_t}$ is a vector of temporal controls including time of day, day of week and month of year, as well as the number of other physicians available in the ED at time t. SystemLoad$_t$ is a measure of the system workload (number of patients currently being processed at time t). Workload$_{it}$ is the number of patients under the care of physician i at time t. The specifications correspond to whether an easy patient is picked up or a difficult patient is picked up. The baseline case is that a difficult case is picked up. A positive value for θ therefore supports our hypothesis that a physician is more likely to pick up an easy patient as workload increases, supporting Hypothesis 1. We use a full information maximum likelihood estimator to obtain the parameters of interest of the choice model given by the equation above.

3.3.2 Short-Term Productivity Effects

We next examine whether picking up an easy patient is in general associated with a short-term improvement in productivity, as measured by the service time. The service time depends on various patient level characteristics. In particular, the case severity of the patient, insurance status, race, gender and age could impact the service processing times. Theory does not postulate a specific functional form for the relationship between length of stay and our patient-level controls. For the purposes of our hypothesis, the exact functional form is unimportant as long as the length of stay is decreasing in the severity level of the patient. Therefore, we propose the

following log-linear specification to examine the relationship between the difficulty level of the patient j (EASY$_j$) and the total service time associated with patient i (SVCTIME$_{ijt}$):

$$\log\left(SVCTIME_{ijt}\right) = \alpha_i + \boldsymbol{X_j}\beta + \boldsymbol{Z_t}\gamma + \delta System_t + \theta_1 Workload_{it} + \theta_2 EASY_j + \varepsilon_{ijt}$$

$\boldsymbol{X_j}$ captures the sources of patient level heterogeneity including race, gender, age, status of admission, and payment status. α_i is the physician fixed effect, which allows us to account for physician level unobserved heterogeneity such as efficiency, motivation, work habits, and general baseline productivity.

In addition, our model also accounts for seasonality at the monthly, weekly, and daily levels, as well as the availability of physicians, indicated by the vector $\boldsymbol{Z_t}$. The variable System$_t$ accounts for system-level workload at time t, and accounts for any system-level congestion, independent of the physician's workload, which could impact the time taken to discharge patient j.

We hypothesize that θ_2 is negative, denoting a faster rate of discharge when the patient is less acute. A negative value of θ_2 would confirm Hypothesis 3 that by picking up easy patients, physicians are able to generate higher service rates (or number of patients processed).

3.4 Results: Task Selection and Short-Term Productivity

We find that physicians are more likely to pick up easier patients (compared to picking up a difficult patient) when the workload is higher. In Table 2, specification (1) excludes the physician fixed effect, and we find that coefficient of physician workload is 0.175 ($p < 0.01$). Specification (2) excludes the temporal sources of seasonality. We find that excluding the temporal heterogeneity still yields the same general result (0.137, $p < 0.01$). Specification (3) excludes the number of physician available. Again we do not observe any impact on our coefficient of interest. Specification (4) is our full model, and accounts for physician fixed effects, and temporal considerations. We find that our estimate for workload is 0.166 ($p < 0.01$). This corresponds to increased odds of 18% in picking up an easier patient when the physician's workload marginally increases by one. As a final test of robustness, we examine specification (5), which includes patient level controls such as age, gender, race, and payer status. These variables are typically not available to the physician prior to patient pickup, and therefore should not be correlated with the pickup decision. We find that the inclusion of these patient controls

has no appreciable impact on our estimate for workload, providing support that our estimates are not subject to confounding. In general, we find that our results are robust to various alternative specifications, attributed to the exogenous nature of patient arrivals into the ED.

We find that the effect of higher levels of physician workload is associated with a shorter length of stay for the patient (Table 3). Specification (1) excludes the physician fixed effect, but includes patient-level and temporal sources of heterogeneity. Specification (5) is our full model, which includes physician fixed effects, temporal considerations, and various patient-level sources of heterogeneity. We find that an easier patient has a shorter length of stay (coefficient of EASY = -0.88, p < 0.01). We find that our estimates are not economically different across the various model specifications, indicating robustness of our results. In conjunction with our results for task selection, Table 3 confirms our third hypothesis that by selecting easier tasks, short-term productivity improves. The interaction term between physician workload and EASY is positive. This suggests that at increased levels of workload, easy patients can increase the length of stay. This interaction effect, however, is small compared to the coefficient of EASY. The coefficient of the interaction term is smaller than the coefficient on EASY in magnitude by a factor of approximately 14 (0.88/0.062). This means that a physician can increase productivity by picking up easy patients up to a level of 14. In practice, the level of physician workload is significantly lower, suggesting that the pickup and discharge of easy patients generally leads to an increase in overall short-term physician productivity.

3.5.1 Physician-Level Heterogeneity in Task Selection

To examine the long-term effects of task selection, we first classify the physicians based on the extent to which they exhibit actions consistent with task completion bias. Our goal is to examine whether physicians who consistently select easier tasks vary systematically in their aggregate level of productivity as compared to physicians who do not consistently select easier tasks. In the prior section, we obtained the average level of easy versus hard task selection across the entire set of physicians in the study. However, the extent of this behavior may vary from physician to physician. To examine the behavior at the individual physician level, we use the following linear probability model specification.[2]

---

[2] We use the LPM here because the interaction coefficients are not impacted by the non-linearity. We also estimate the model using a logistic specification, and find that the results are similar.

$$EASY_{ijt} = \alpha_i + X_j\beta + Z_t\gamma + \delta SystemLoad_t + \theta Workload_{it} + \alpha_i * Workload_{it} + \varepsilon_{ijt}$$

In this specification, the interaction of the physician fixed effect (α_i) with the level of physician workload allows us to account for task selection heterogeneity at the individual physician level. Specifically, for each physician, we obtain a numerical estimate for the extent of easy task selection in response to increased workload; this estimate is then utilized to rank-order the physicians based on their tendencies to exhibit this behavior.

Our estimate of the model demonstrates that physicians exhibit significant heterogeneity in task selection behavior. At the individual physician level, after accounting for the physician heterogeneity in task selection behavior, the baseline effect of selecting easy tasks across all physicians (indicated by the coefficient θ above) is 0.021 (p < 0.01). However, there is significant heterogeneity in workload-induced task selection behavior across the physicians. Specifically, the interaction term ($\alpha_i * Workload_{it}$) for the set of physicians in the sample has a mean of -0.017 and standard deviation of 0.018, while the median is -0.021 (p < 0.01). Once we estimate the task selection behavior heterogeneity at the physician level, we classify physicians into two categories following a median split based on the coefficient estimate for the obtained interaction term ($\alpha_i * Workload_{it}$). Specifically we denote TCB_HI_i = 1 if physician i has a higher tendency to exhibit TCB, and TCB_HI_i = 0 otherwise.

3.5.2 Long-Term Productivity Effects

To examine the long-term impact of TCB on productivity, we estimate the following model:

$$\log(SVCTIME_{ijt}) = \mu\, TCB\_HI_i + X_j\beta + Z_t\gamma + \delta System_t + \theta_1 Workload_{it} + \varepsilon_{ijt}$$

In the specification above, a positive value for μ would confirm our hypothesis that physicians who are more likely to exhibit TCB take longer to process patients on average, and therefore have lower long-term productivity than physicians who do not exhibit TCB.

To examine the long-term revenue implications of TCB, we consider the following specification, where we examine the effect of TCB on the revenue as follows:

$$RVU_{ijt} = \mu\, TCB\_HI_i + X_j\beta + Z_t\gamma + \delta System_t + \theta_1 Workload_{it} + \varepsilon_{ijt}$$

A negative value for μ would confirm our hypothesis that physicians who exhibit TCB generate lower revenues for the hospital.

3.6 Results: TCB and Long-Term Productivity

Table 4 presents the results for the effect of TCB on long-term productivity as measured by the service time. We find that the coefficient of TCB is positive, indicating that physicians that exhibit TCB have lower processing rate, and therefore lower productivity. This finding is robust to various alternative specifications that include temporal seasonality (specification 1), patient level-controls (specification 2), and both temporal and patient controls (specification 3). Specification (4) is our complete specification that includes all observed sources of patient-level and temporal heterogeneity, as well as the system load. We find that coefficient for TCB is 0.056 ($p < 0.01$). This suggests that physicians who exhibit TCB have lower overall productivity by 5.6% on average, after accounting for task difficulty.

Table 5 presents the results for the effect of TCB on RVU's. Our results are consistent with the findings on productivity. The coefficient estimates for TCB across all of the model specifications, which include differing levels of sources of heterogeneity, are negative and statistically significant. Namely, physicians who tend to exhibit high TCB also tend to generate lower RVU's for the hospital, indicating reduced profitability for the hospital. From specification (5), which is our complete model, we find that physicians who exhibit TCB generate -0.291 units fewer RVU's per patient.

Collectively, these results indicate that in the short run, TCB appears to increase productivity. We see that this is attributed primarily to the fact that easy tasks take a shorter amount of time to complete. In the long run, the tendency for TCB leads to fewer difficult tasks being selected. In other words, TCB amounts to "kicking the can down the road" at the detriment of long-term productivity.

4. Study 2: Lab Studies to Conceptually Replicate and Identify Mechanisms

Having seen evidence in the field, of how individuals select easy tasks over hard tasks when workload is increasing, we now turn to the lab to both replicate conceptually the finding and then to examine the potential mechanism. As a first step we conducted two pilot studies to see if we had evidence for a focus on easier tasks.

4.1 Pilot Study 1

We recruited 50 students (34% men). We presented them with 3 different task descriptions and an example of each task in random order and asked them to rate how hard or easy they found each task on a 7-point scale (1 = very easy, to 7 = very hard).

In the first task (typing task), participants were presented with the picture of sentences on the screen and were asked to type the sentence in the space provided. In the second task (math task), participants were presented with equations and were asked to type in the answers to each equation one at a time into the space provided (e.g., 83-45=?). The third task (number recognition task) was a number recognition task in which participants were provided with a string of numbers and were asked to identify whether there was a "4" among the numbers or not (e.g., 3846778).

Afterwards, we randomly assigned each participant to one of the two conditions (busy vs. control). In the control condition, participants were asked to choose one of the tasks to complete right now for 2 minutes. In the busy condition, we told them to imagine they are very busy right now and then asked them to choose one of the tasks to complete right now for 2 minutes.

4.1.1 Pilot Study 1 Results

Running paired sample t-tests, participants rated the number recognition task (M = 2.36, SD = 1.35) easier than both the typing task (M = 3.28, SD = 1.51), t(49) = 3.56, p = .001 and the math task (M = 3.30, SD = 1.36), t(49) = 4.20, p < .001. There was no significant difference in perceptions of how hard tasks were between the typing and math tasks (p = .94).

We tested for the differences between the two conditions for the percentage of people who chose the easier task (the recognition task) as opposed to the harder tasks (either the math or typing task). As predicted in Hypothesis 1, a higher percentage of participants chose the easy task in the busy condition (52.0%, 13/25) than in the control condition (24.0%, 6/25), $\chi 2$ = 4.16, p = .041.

4.2 Pilot Study 2

We recruited 88 students (48.9% men). Participants were randomly assigned to one of the two conditions (busy vs. control).

In the busy condition, we told them to imagine they are very busy right now, that is to put themselves in the shoes a person that is very busy at the moment. All participants were informed that they could choose one of two tasks to complete next, that the choice is entirely up to them.

To give them a better sense, they were given instructions and examples of each task. We used the number recognition and typing tasks from the pilot study 1.

Afterwards, participants chose one of the two tasks. We then presented them with their task of choice to complete for three minutes.

4.2.1 Pilot Study 2 Results

We tested for the differences between the two conditions for the percentage of people who chose the easier task (the number recognition task) as opposed to the harder tasks (the typing task). As predicted in Hypothesis 1, a higher percentage of participants chose the easy task (number recognition task) in the busy condition (81.4%, 35/43) than in the control condition (60.0%, 27/45), $\chi 2 = 4.84$, p = .028.

4.3 Study 2: Identifying the Mechanism

Our second hypothesis suggested that individuals focus on easier tasks because: (a) a sense of progress and (b) completing a task makes people feel good. We tested this hypothesis in the laboratory 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 those who had experienced a higher workload in the first part of the study would feel as if they had accomplished more and thus would choose easier tasks either to maintain the sense of progress or to continue to feel good.

4.3.1 Participants. Two-hundred thirty-two participants recruited from the subject pool of a university in the northeastern United States (56% male) participated in the study in exchange for a $20 payment. We calculated our sample size based on an estimate of an effect size $d=.4$, requiring a sample size of approximately 200 participants for a study powered at 80%. We ended up with a bigger sample due to the number of participants who showed up for the sessions compared to what we had scheduled.

We randomly assigned participants to one of two conditions: high-workload and low-workload.

4.3.2 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 in on the screen, in the appropriate window, as much of the text as they could. We manipulated participants' workload by asking half of them (high workload condition)

to listen to a song while working on 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 (Moore and Healy 2008). The other half (low workload condition), instead, 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.

4.3.3 Procedure. Participants were told that the study consisted of different tasks and short questionnaires. They first completed the first task, which consisted in transcribing a text based on a picture of it. They saw the picture of the text on the screen and were asked to copy it in the appropriate space on the screen, by typing it. Participants had only three minutes to finish working on the task. 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."

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 two measures, perceived sense of progress (three items) and good feelings about task completion (three items). The items used to measure sense of progress were: 1) I am feeling 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 ($\alpha = .97$). The items used to measure good feelings about task completion were: 1) I feel good about having completed the task, 2) I am happy about what I have accomplished so far, and 3) I feel satisfied with the work I completed. We averaged participants' answers across these three items to create a measure of feeling good about task completion ($\alpha = .96$).

Next, participants moved on to the next task, called "Letter Processing and Word Creation." The task is focused on demonstrating how people process letters in order to form words. The instructions informed participants that, "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 difficult.
Version 2: Most people find this version of the task to be not that challenging but rather easy.
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 onto 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 common English words they were able generate.

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, 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 manipulation check (α = .88).

Finally, participants answered a few demographic questions.

4.3.4 Results

Table 6 shows the descriptive statistics of the variables captured in Study 2 by condition.

4.3.4.1 Manipulation check. As expected, participants reported feeling as if they had a higher workload in the high-workload condition than in the low-workload condition, $t(230) = 2.18$, p = .03. They also transcribed fewer words of the text in the high-workload condition than in the low-workload condition, $t(230) = -2.44$, p = .015. These results suggest that our manipulation was effective.

4.3.4.2 Perceived progress and feelings about task completion. Participants in the high-workload condition reported perceiving a lower sense of progress and feeling less good about task completion as compared to participants in the low-workload condition, $t(230) = 5.13$, p < .001 and $t(230) = 4.24$, p < .001, respectively.

4.3.4.3 Choosing the difficult task. As predicted in Hypothesis 1, a lower percentage of participants chose the difficult task in the high-workload condition (14.4%, 16/111) than in the low-workload condition (38.8%, 47/121), $\chi2$ (1, N = 232) = 17.47, *p* < .001.

4.3.4.4 Performance on the second task. Independent of the choice they 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, $t(230) = .46$, $p = .65$.

4.3.4.5 Mediation analyses. Our second hypothesis predicted that individual's task selection would be driven by either (H2a) feelings of progress or (H2b) positive feelings from task completion. We tested perceived progress as a potential mediator and did not find evidence in support of this mechanism as the one explaining our effects. We did find support for H2b, however.

We tested whether feeling good about task completion mediated the relationship between our workload conditions and the choice of difficult task, using the bootstrapping approach outlined by Preacher and Hayes (2004). Based on bootstrapping (with 10,000 iterations), we estimated the direct and indirect effects of the high-workload condition via our proposed mediator on our dependent variable: the choice to complete a difficult task as the second task. Our manipulation had a significant effect on feeling good about task completion (as shown by the analyses above), which, in turn, significantly affected the choice of difficult task ($B = .20$ S.E. = .09, $p = .021$). In contrast, the effect of our manipulation was reduced (from $B = 1.33$, S.E. = .33, $p < .001$ to $B = 1.16$, S.E. = .34, $p = .001$) when feeling good about task completion after the first task was included in the equation. The 95 percent bias-corrected confidence interval for the size of the indirect effect excluded zero (.03, .37), suggesting significant mediation.

4.3.5 Discussion

The results of Study 2 not only conceptually replicate our first hypothesis, but also they show that, as we predicted in our second hypothesis, feeling good about task completion on the first task would explain why a high workload leads to the choice of easy versus difficult tasks afterwards.

5. Discussion & Conclusion

In this paper we look to understand how individuals manage tasks from a common queue. We investigate the short-term and long-term performance implications. In so doing we make several contributions to the literature. First, using both the field and the lab we provide evidence of a task completion bias. As individuals experience higher levels of load they select easier tasks. This finding is important as it provides another 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 that has examined discretionary effort allocation – such as the speed - quality tradeoff (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 not only study task selection in more detail, but also how task selection and the speed-quality tradeoff may interact. Analytical work should investigate how the task completion bias 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 showing that task completion bias can create system benefits, at least in the short-term. In our main models we show that when workers select easier tasks they complete the work faster. One can also explore the interaction between workload and selecting easier tasks. Examining the coefficient on Physician Workload x Easy in Table 5, we see that the coefficient is positive – suggesting that as the workload increases the benefits from easy tasks decrease. Although it continues to have a positive impact on short-term performance there are limits as there are diminishing returns to the gain.

In addition to examining how selecting easier tasks impacts completion time, it is possible to explore how task completion bias impacts variability in completion times. To examine this effect, we first obtain the coefficient of variation of the task completion times for each physician. To account for seasonality, we further group the tasks completed according to the month of year, day of week, and time of day. This yields the measure $CV(SVCTIME_{it})$, which is the coefficient of variation in service time for all of physician i's patients treated in a given month in a given day during a given time of day. To examine whether the coefficient of

variation differs across physicians who exhibit the task completion bias, compared to those who do not (again using a median split), we estimate the following model:

$$CV(SVCTIME_{it}) = \rho\, TCB\_HI_i + \boldsymbol{Z_t}\gamma + \varepsilon_{ijt}$$

Our goal is to estimate the coefficient ρ. A negative value for ρ indicates that physicians who exhibit TCB have lower overall variability in their service times. Z_t includes temporal fixed effect, including month, day of week, and time of day. Table 7 shows the results of this regression. We estimate ρ to be negative (Table 7, TCB = -6.025, p < 0.01), confirming that physicians who exhibit a higher degree of TCB interestingly have a lower coefficient of variation. As such, by picking up easy tasks rather than difficult tasks when workload increases, workers can prevent overall task completion times from significantly increasing during periods of high workload, thus reducing the overall variability in task completion times.

Our third contribution investigates the longer term impact of the task completion bias. Although prior work suggests that improved performance may eventually hinder performance due to overwork (KC and Terwiesch 2009; Staats and Gino 2012; Kuntz et al. 2015) here we show a negative longer-term effect that is consistent with a lack of learning 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. Although always selecting the harder task may be suboptimal, if one continually chooses the exploitation path then longer term performance suffers. More generally, time is a key dimension to consider in the operational construction of a system. Just as prior work shows different benefits costs in the short and long term to factors such as task variety (Staats and Gino 2012) or process compliance (Staats et al. 2017), here we do so with task selection. More work is needed to understand time and its consequences.

Finally, in our work we are the first to show the mechanism through which completion benefits performance – the positive feelings that accrue as work is finished. 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 mechanism through which it

occurs, allows us to design systems around a positive mechanism (or avoid a negative one). Here our finding would suggest understanding ways to get the 'completion high' – for example by breaking tasks into smaller pieces – 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, in the field setting of the hospital we are not able to test our mechanism. The lab provides an opportunity to gain more precise control for testing purposes. However, future work should seek to study both our identified mechanism, as well as others, in the field.

Third, our measure of long-term performance is somewhat coarse – as we look at a median split of those that follow the task completion bias. Ideally future work could implement a longer-term field experiment or another approach to study the topic in a more fine-grained manner. Finally, we rely on students as laboratory subjects in the research lab to identify our mechanism. This is an established and valid practice when studying human decision making since we are interested in a general behavior of engaging in the task completion bias. Still, future work could consider creative ways to engage real participants – for example, as Tucker (2015) has done when she runs experiments on practicing nurses at a convention. Fundamentally, our study relies on triangulation for its validity, but the more work – both with respect to the overall finding and with respect to the mechanism – that can be done in the field with real practitioners, the more benefits that will accrue to 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 can accrue when workers have discretion with which to manage their tasks (Bowman 1963; van Donselaar et al. 2010; Ibañez et al. 2017). 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 bias we find a way that people may work against their long-term self-interest.

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, educate workers about task completion bias as a means to address it. In addition, given the performance implications – managers should educate workers about the performance consequences 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 happiness to structure work effectively. As noted, this could mean taking harder tasks and breaking them into sub-components, each with a clearly defined indicator of completion, so that completion occurs more often. Since Taylor's Scientific Management the structure of work has been a primary focus of the operations' field. That is no less true today than it was then – if anything it has become even more of a focus because managers today have more degrees of freedom to structure work – sending different tasks to different people, breaking up the work into different pieces, etc. That means both 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. Increasingly that work is done by individuals and so understanding People Operations is necessary for advancements in both theory and practice. In this paper we draw upon research on workload, individual discretion, and behavioral decision making to understand how workload and task selection interact. We theorize and test in an emergency department setting that under conditions of increased workload individuals may choose to complete easier tasks in order to manage their load. We call this behavior task completion bias. We then investigate the performance implications of task completion bias and find that it aids short-term performance, but harms long-term performance. In addition, turning to the lab we conceptually replicate the task selection effect and show that it occurs due to the positive feelings individuals get from task completion. With these findings we identify an alternative mechanism for the workload-speedup effect previously found in the literature. Altogether, by better structuring work it is possible to aid individual development and organizational performance.

6. References

Amabile, T. and S. Kramer (2011). <u>The progress principle: Using small wins to ignite joy, engagement, and creativity at work</u>. Boston, Harvard Business Press.

Amar, M., D. Ariely, S. Ayal, C. E. Cryder and S. I. Rick (2011). "Winning the battle but losing the war: The psychology of debt management." *Journal of Marketing Research* **48**(SPL): S38-S50.

Anand, K. S., M. F. Paç and S. Veeraraghavan (2011). "Quality–Speed Conundrum: Trade-offs in Customer-Intensive Services." *Management Science* **57**(1): 40-56.

Arlotto, A., S. E. Chick and N. Gans (2014). "Optimal hiring and retention policies for heterogeneous workers who learn." *Management Science* **60**(1): 110 - 129.

Bassamboo, A. and R. S. Randhawa (2015). "Scheduling homogeneous impatient customers." *Management Science*.

Batt, R. J. and C. Terwiesch (2016). "Early task initiation and other load-adaptive mechanisms in the emergency department." *Management Science*.

Bendoly, E., K. Donohue and K. L. Schultz (2006). "Behavior in operations management: Assessing recent findings and revisiting old assumptions." *Journal of Operations Management* **24**(6): 737-752.

Berry Jaeker, J. and A. L. Tucker (2015). "Past the point of speeding up: The negative effects of workload saturation on efficiency and quality." *Management Science*.

Bolton, G. E., A. Ockenfels and U. W. Thonemann (2012). "Managers and students as newsvendors." *Management Science* **58**(12): 2225-2233.

Bowman, E. H. (1963). "Consistency and optimality in managerial decision making." *Management Science* **9**(2): 310-321.

Buell, R. W., T. Kim and C.-J. Tsay (2016). "Creating reciprocal value through operational transparency." *Management Science*.

Cachon, G. and C. Terwiesch (2009). <u>Matching Supply with Demand</u>. Boston, McGraw-Hill Irwin.

Campbell, D. and F. X. Frei (2011). "Market heterogeneity and local capacity decisions in services." *Manufacturing & Service Operations Management* **13**(1): 2-19.

Chan Jr., D. C. (2015). The efficiency of slacking off: Evidence from the emergency department, National Bureau of Economic Research.

Crabill, T. B. (1972). "Optimal control of a service facility with variable exponential service times and constant arrival rate." *Management Science* **18**(9): 560-566.

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

Debo, L. G., L. B. Toktay and L. N. Van Wassenhove (2008). "Queuing for expert services." *Management Science* **54**(8): 1497-1512.

Deo, S. and A. Jain (2015). "Slow first, fast later: Empirical evidence of speed-up in service episodes of finite duration."

Freeman, M., N. Savva and S. Scholtes (2016). "Gatekeepers at work: an empirical analysis of a maternity unit." *Management Science*.

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

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

Gino, F. and G. P. Pisano (2008). "Toward a theory of behavioral operations." *Manufacturing and Service Operations Management* **10**(4): 676-691.

Green, L., N. Savva and S. Savin (2012). "Nursevendor problem": Personnel staffng in the presence of endogenous absenteeism.

Green, L. V., S. Savin and B. Wang (2006). "Managing patient service in a diagnostic medical facility." *Operations Research* **54**(1): 11-25.

Hayes, R. H. (2002). "Wick Skinner: A life sailing against the wind." *Production and Operations Management* **11**(1): 1-8.

Heath, C., R. P. Larrick and G. Wu (1999). "Goals as reference points." *Cognitive psychology* **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." *The Journal of General Psychology* **97**(1): 17-71.

Hopp, W. J., S. M. R. Iravani and G. Y. Yuen (2007). "Operations systems with discretionary task completion." *Management Science* **53**(1): 61-77.

Huckman, R. S., B. R. Staats and D. M. Upton (2009). "Team familiarity, role experience, and performance: Evidence from Indian software services." *Management Science* **55**(1): 85-100.

Ibañez, M., J. Clark, R. S. Huckman and B. R. Staats (2017). "Discretionary Task Ordering: Queue Management in Radiological Services." *Management Science*.

KC, D. and B. R. Staats (2012). "Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance." *Manufacturing & Service Operations Management* **14**(4): 618-633.

KC, D., B. R. Staats and F. Gino (2013). "Learning from my successes and others' failure: Evidence from minimally invasive cardiac surgery." *Management Science* **59**(11): 2435-2449.

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

Kim, S.-H., C. W. Chan, M. Olivares and G. Escobar (2015). "ICU admission control: An empirical study of capacity allocation and Its implication for patient outcomes." *Management Science* **61**(1): 19-38.

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

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

Kuntz, L., R. Mennicken and S. Scholtes (2015). "Stress on the Ward: Evidence of Safety Tipping Points in Hospitals." *Management Science* **61**(4): 754-771.

Lapré, M. A. and I. M. Nembhard (2010). "Inside the organizational learning curve." *Foundations and Trends in Technology, Information and Operations Management* **4**(1): 1-103.

March, J. G. (1991). "Exploration and exploitation in organizational learning." *Organization Science* **2**(1): 71-87.

Moore, D. A. and P. J. Healy (2008). "The trouble with overconfidence." *Psychological Review* **115**(2): 502-517.

Narayanan, S., S. Balasubramanian and J. M. Swaminathan (2009). "A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment." *Management Science* **55**(11): 1861-1876.

Newell, A. and H. A. Simon (1972). Human problem solving, Prentice-Hall Englewood Cliffs, NJ.

Phillips, R., A. S. Şimşek and G. v. Ryzin (2015). "The effectiveness of field price discretion: Empirical evidence from auto lending." *Management Science*.

Pinedo, M. L. (2012). Scheduling: Theory, Algorithms, and Systems. New York, Springer.

Powell, A., S. Savin and N. Savva (2012). "Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient." *Manufacturing & Service Operations Management*.

Saghafian, S., W. J. Hopp, M. P. Van Oyen, J. S. Desmond and S. L. Kronick (2014). "Complexity-augmented triage: A tool for improving patient safety and operational efficiency." *Manufacturing & Service Operations Management* **16**(3): 329-345.

Schultz, K. L., D. C. Juran and J. W. Boudreau (1999). "The effects of low inventory on the development of productivity norms." *Management Science* **45**(12): 1664-1678.

Schultz, K. L., D. C. Juran, J. W. Boudreau, J. O. McClain and L. J. Thomas (1998). "Modeling and worker motivation in JIT production systems." *Management Science* **44**(12): 1595-1607.

Schweitzer, M. E. and G. P. Cachon (2000). "Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence." *Management Science* **46**(3): 404-420.

Shumsky, R. A. and E. J. Pinker (2003). "Gatekeepers and referrals in services." *Management Science* **49**(7): 839-856.

Smiddy, H. and L. Naum (1954). "Evolution of a 'Science of Managing' in America." *Management Science* **1**(1): 1-31.

Song, H., A. Tucker and K. Morrell (2015). "The diseconomies of queue pooling: An empirical investigation of emergency department length of stay." *Management Science*.

Staats, B. R., H. Dai, D. A. Hofmann and K. L. Milkman (2017). "Process compliance and electronic monitoring: Empirical evidence from hand hygiene in healthcare." *Management Science* **63**(5): 1563-1585.

Staats, B. R. and F. Gino (2012). "Specialization and variety in repetitive tasks: Evidence from a Japanese bank." *Management Science* **58**(6): 1141-1159.

Staats, B. R., D. KC and F. Gino (2016). "Maintaining beliefs in the face of negative news: The moderating role of experience." *Management Science*.

Staw, B. M., L. E. Sandelands and J. E. Dutton (1981). "Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis." *Administrative Science Quarterly* **26**(4): 501-524.

Stidham Jr, S. and R. R. Weber (1989). "Monotonic and insensitive optimal policies for control of queues with undiscounted costs." *Operations Research* **37**(4): 611-625.

Tan, T. F. and S. Netessine (2014). "When does the devil make work? An empirical study of the impact of workload on server's performance." *Management Science* **60**(6): 1574-1593.

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

Terwiesch, C. (2017). "MSOM Fellows Talk."

Tucker, A. L. (2007). "An empirical study of system improvement by frontline employees in hospital units." *Manufacturing & Service Operations Management* **9**(4): 492–505.

Tucker, A. L. (2015). "The impact of workaround difficulty on frontline employees' response to operational failures: A laboratory experiment on medication administration." *Management Science*.

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

Weick, K. E. (1984). "Small wins: Redefining the scale of social problems." *American Psychologist* **39**(1): 40-49.

Zipkin, P. H. (2000). Foundations of inventory management, McGraw-Hill New York.

7. Tables

Table 1. Summary Statistics

	mean	median	sd	1	2	3	4	5	6	7	8
1.Easy	0.178	0.000	0.383								
2.Service Time	3.203	2.550	2.575	-0.13							
3.RVU	5.150	1.418	7.164	-0.17	0.16						
4.Age	38.159	36.000	22.784	0.12	0.18	-0.04					
5.Female	0.510	1.000	0.500	-0.06	0.04	0.02	0.04				
6.Medicaid	0.176	0.000	0.381	-0.10	-0.02	0.03	-0.16	0.03			
7.Num. of Physicians	5.238	5.000	2.355	-0.25	-0.06	0.02	-0.24	0.05	0.13		
8.System Load	15.511	15.000	7.171	-0.07	0.05	0.01	-0.17	0.04	0.08	0.62	
9.Physician Workload	4.296	4.000	2.598	0.13	-0.06	-0.06	-0.05	-0.00	-0.00	0.04	0.37

note: All the correlations are statistically significant at the 5% level

Table 2. Task Selection Bias

	(1)	(2)	(3)	(4)	(5)
	coef/se	coef/se	coef/se	coef/se	coef/se
Physician Fixed Effects		x	x	x	x
Temporal Fixed Effects	x		x	x	x
Patient Controls					x
Physicians Available	-0.626***	0.081		-0.136*	-0.140**
	(0.082)	(0.095)		(0.071)	(0.069)
System Load	-0.025***	0.048***	-0.008	-0.001	0.002
	(0.007)	(0.013)	(0.011)	(0.013)	(0.013)
Physician Workload	0.175***	0.137***	0.173***	0.166***	0.160***
	(0.028)	(0.035)	(0.039)	(0.041)	(0.041)
Number of observations	93,797	92,033	92,033	92,033	92,033
Log-Likelihood	-30,513.14	-15,576.83	-14,690.31	-14,658.25	-14,491.09

note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Patient controls include age, gender, race, insurance, source of arrival and severity

Table 3. Temporary Speedup Effects

	(1)	(2)	(3)	(4)	(5)
	coef/se	coef/se	coef/se	coef/se	coef/se
Physician Fixed Effects		x	x	x	x
Temporal Fixed Effects	x		x	x	x
Patient Controls	x	x		x	x
Physicians Available	-0.068***	0.006*	0.003	0.001	0.003
	(0.011)	(0.003)	(0.004)	(0.004)	(0.004)
System Load	0.023***	0.026***	0.025***	0.025***	0.025***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Physician Workload	-0.022***	-0.002	-0.002	0.010	-0.001
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Easy	-0.907***	-0.877***	-0.902***	-0.593***	-0.880***
	(0.095)	(0.069)	(0.072)	(0.056)	(0.071)
Physician Workload x Easy	0.095***	0.065***	0.062***		0.062***
	(0.006)	(0.010)	(0.009)		(0.009)
Number of observations	93,675	93,675	93,675	93,675	93,675
Log-Likelihood	-98,286.61	-88,594.21	-88,988.14	-88,690.44	-88,281.14

note: *** p<0.01, ** p<0.05, * p<0.1. Patient controls include age, gender, race, insurance, source of arrival and severity

Table 4. Long Term Productivity Effects

	(1)	(2)	(3)	(4)
	coef/se	coef/se	coef/se	coef/se
Temporal Fixed Effects	x		x	x
Patient Controls		x	x	x
Easy		-0.396***	-0.446***	-0.452***
		(0.006)	(0.007)	(0.007)
Physicians Available	-0.046***	-0.065***	-0.039***	-0.071***
	(0.002)	(0.001)	(0.001)	(0.002)
System Load	0.024***	0.028***		0.024***
	(0.001)	(0.000)		(0.000)
Physician Workload	-0.014***	-0.010***		-0.008***
	(0.001)	(0.001)		(0.001)
TCB	0.030***	0.094***	0.044***	0.056***
	(0.005)	(0.005)	(0.005)	(0.005)
Number of observations	93,675	93,675	93,675	93,675
Log-Likelihood	-101,362.12	-100,089.37	-100,266.17	-99,027.70

note: *** p<0.01, ** p<0.05, * p<0.1. Patient controls include age, gender, race, insurance, source of arrival and severity

Table 5. Long Term Revenue Effects

	(1)	(2)	(3)	(4)	(5)
	coef/se	coef/se	coef/se	coef/se	coef/se
Temporal Fixed Effects	x		x	x	x
Patient Controls		x	x	x	x
Easy					-3.059***
					(0.070)
Physicians Available	0.005	-0.041***	0.055***		
	(0.016)	(0.015)	(0.015)		
System Load	0.034***	0.038***		0.033***	0.006
	(0.005)	(0.005)		(0.005)	(0.005)
Physician Workload	-0.195***	-0.204***		-0.194***	-0.127***
	(0.011)	(0.011)		(0.011)	(0.011)
TCB	-0.560***	-0.633***	-0.642***	-0.542***	-0.291***
	(0.058)	(0.057)	(0.058)	(0.058)	(0.057)
Number of observations	80,651	80,651	80,651	80,651	80,651
Log-Likelihood	-272,729.28	-272,846.58	-272,862.10	-272,699.51	-271,743.18

note: *** p<0.01, ** p<0.05, * p<0.1. Patient controls include age, gender, race, insurance, source of arrival and severity

Table 6. Descriptive statistics of the variables captured in Study 2 by condition

	High workload condition	Low workload condition
Perceived workload (man check)	5.83 (1.25)	5.44 (1.44)
Number of words copied	83.89 (35.67)	72.73 (33.82)
Perceived sense of progress	3.20 (1.65)	4.33 (1.68)
Feelings about task completion	3.26 (1.74)	4.25 (1.82)
Choice of difficult task	14.4%	38.8%
Performance score on task #2	75.95 (41.87)	78.43 (41.02)

Table 7. Task Completing Biased Physicians Have Lower Variability In Service Time

	(1)	(2)
	coef/se	coef/se
Time Fixed Effects		x
TCB	-6.042***	-6.025***
	(0.789)	(0.787)
Number of observations	4,128	4,128
Log-Likelihood	-18,944.12	-18,923.48

note: *** p<0.01, ** p<0.05, * p<0.1. Time fixed effects account for month, day of week, and time during the day