Task Variety in Professional Service Work: When It Helps and When It Hurts

Emmanouil Avgerinos

IE Business School, IE University, Maria De Molina 12, 28006, Madrid, Spain

emmanouil.avgerinos@ie.edu

Bilal Gokpinar*

UCL School of Management, University College London, 1 Canada Square, London, E14 5AB, UK
b.gokpinar@ucl.ac.uk

In a wide range of professional service firms, individuals perform a variety of tasks which are highly cognitive and knowledge intensive yet repetitive in nature, providing significant opportunities for learning. In addition, individuals in such environments tend to enjoy considerable discretion in managing when and how they perform their tasks. In light of these observations, we investigate task allocation and timing strategies that may enhance or inhibit learning and productivity for professional service workers. Specifically, we focus on the role of task variety. We use a detailed dataset of 3,275 coronary artery bypass surgeries in a private European hospital over seven years to examine the effect of concurrent and non-concurrent exposure to task variety on learning and productivity on a focal task. We find that while concurrent exposure to variety has a positive impact on focal productivity, non-concurrent exposure to variety has a negative impact on it. Our results also suggest that short-term exposure to variety amplifies these relationships.

Key words: task variety, learning, productivity, task timing.

History: Received: August 2014; accepted March 2018 by Aleda Roth after four revisions.

1. Introduction

Professional service firms globally generate annual sales over $3 trillion and represent 7-8% of total service sector revenue in advanced economies (McKinsey 2012). This percentage is even higher in service-based economies such as Britain, where 15% of the GDP and 14% of employment comes from professional services firms (PwC 2012). Following Von Nordenflycht's (2010) characterization, the professional service industry broadly includes accounting, advertising and marketing, management consulting, architecture, legal services, scientific research services, and physician practices. While these firms have distinct characteristics at a high level, including knowledge intensity, low capital intensity, and a professionalized workforce (Von Nordenflycht 2010), their operations are also distinguished by several key features that relate to how their employees perform their work.

First, the majority of the work that professional service workers perform is quite repetitive in nature. For example, management consultants follow similar steps in their engagement with clients, from initiation to contracting and final deliverables; legal professionals draft legal documents by engaging in

---

* Corresponding author

a similar set of activities; and surgeons follow a certain set of procedures in the operating room. While most activities and tasks may be quite similar from one job to another, workers still engage in high levels of cognitive activity, presumably due to variation in work content across tasks (e.g., differences between consulting projects, surgeries, or legal cases). Hopp et al. (2009) consider this key observation in their classification of white-collar work: they call, for instance, consulting and legal services intellectual and routine work. As a result of the repetitive nature of the work and the significant opportunities for learning that these settings offer, individuals inevitably transfer their past experience and knowledge when they work on subsequent tasks (Tversky 1977, Gick and Holyoak 1987, Zollo and Reuer 2010).

A second important feature of professional service work is that individuals tend to have a relatively high degree of discretion in managing when and how they perform their tasks (Hopp et al. 2007, Ibanez et al. 2017). Compared with workers in other professions, they enjoy more control and flexibility over decisions regarding task sequences, including whether to perform certain tasks concurrently or individually and in smaller pieces or larger chunks.

Considering these two key features of professional service work—first, its repetitive nature and many learning opportunities across tasks and, second, workers' potential discretion in managing when and how to perform various tasks—raises an important operational question: How should professional service workers perform their various tasks to achieve greater learning and productivity over time? More specifically, in performing the variety of tasks that a professional service worker is supposed to carry out, are there certain task-timing configurations that enhance or inhibit learning and productivity?

Our study addresses these questions by focusing on how workers perform various tasks. We distinguish between concurrent and non-concurrent exposure to variety and examine productivity implications of these two approaches to organizing work. *Concurrent variety* refers to performing another task concurrently with the focal one, whereas *non-concurrent variety* refers to performing another task independently (i.e., at a different time) from the focal one. Because many professional service workers inevitably perform a variety of tasks, we seek to understand when and how exposure to variety helps, and when and how it hurts individuals' focal productivity.

Exposure to variety through successful knowledge transfer to the focal task may have a positive impact on performance (Schilling et al. 2003, Boh et al. 2007, Staats and Gino 2012), but too much exposure to variety may be detrimental to productivity (Narayanan et al. 2009). Also, task variety could be confusing for individuals (Allport et al. 1994) and therefore decrease their subsequent focal productivity due to switching costs and warm-up periods (Cellier and Eyrolle 1992, Monsell 2003). Like the studies that have produced these findings, our study explores the productivity implications of exposure to task variety, but with an important distinction. We propose that the influence of task variety on productivity critically depends on *when* other tasks are performed in relation to the focal task. We suggest that knowledge transfer and learning mechanisms are quite different when other tasks are

performed concurrently vs. non-concurrently with the focal task, which leads to differentiated and, in fact, contrasting effects on productivity.

We develop and test four hypotheses regarding professional service workers' concurrent and non-concurrent exposure to variety by examining their effect on productivity in subsequent focal tasks. We find that concurrent exposure to variety enhances productivity, whereas non-concurrent exposure to variety is detrimental to productivity. Non-concurrent variety reduces productivity by generating inaccurate transfers and mapping processes between tasks performed at different times that are similar on the surface but structurally different (Gick and Holyoak 1987). Surface similarities may include similar contexts (Craik and Tulving 1975), goals, and processes (Bransford and Franks 1976, Tulving and Thomson 1973). In contrast, structural similarities are deeper resemblances between the elements and underlying causal structure of tasks (Gick and Holyoak 1987); structural similarities enable workers to derive causal inferences from one task to apply to another task (Holyoak and Koh 1987). With concurrent variety, surface similarities do not mask structural differences, because the worker is performing the tasks at the same time. In fact, concurrent performance is highly conducive to learning: it enables 'implicit learning' (Reber 1989, Wulf and Schmidt 1997) and facilitates cognitive skill acquisition through the discrimination process (Anderson 1982), all of which leads to better comprehension of the focal task and increased productivity.

In addition, our results suggest that short-term exposure to variety amplifies the influence of long-term exposure on subsequent focal productivity. That is, recent (short-term) concurrent variety increases the positive influence of concurrent variety on focal productivity, whereas recent non-concurrent exposure to variety strengthens the negative impact of non-concurrent variety on focal productivity.

As we test our hypotheses and explore how professional service workers can achieve higher productivity by better organizing the variety of tasks they perform, we want to emphasize an important empirical consideration. Because many professional service workers have significant discretion over when and how to perform their tasks, there may be inherent endogeneity in many professional service settings, which could pose problems in empirical identification. The ideal test of our hypotheses requires a professional service setting where decisions about which tasks to perform, when to perform different tasks, and whether to perform them concurrently or non-concurrently are not up to individuals' discretion but are instead determined exogenously. We test our hypotheses using a detailed dataset of 3,275 coronary artery bypass graft (CABG) operations from the cardiac unit of a private European hospital over seven years. Because a patient's need is the primary driver of the type, nature, and timing of an operation and these are not up to the discretion of the surgeon, we believe that this is an ideal setting in which to investigate the effects of task variety on productivity.

Our paper offers a number of significant contributions to the service operations management literature. First, by focusing on an under-studied service sector, namely operations of professional services firms (Roth and Menor 2003, Lewis and Brown 2012), our study examines how different task allocation and composition strategies may influence learning and subsequent focal productivity of

professional service workers. This, we believe, is also a step towards answering Argote et al.'s (2003) call for more research to identify "mechanisms and conditions under which experience is beneficial (or harmful) for learning outcomes" and a step towards answering their question about whether different types of experience may provide better understanding of the task (p. 579). Second, our paper contributes to the growing literature on task variety and productivity by introducing a new dimension of variety which has not been considered before: concurrent or non-concurrent exposure. In a similar study, Staats and Gino (2012) focused on a single current task, studied when different tasks took place (on the same day or in the past) with respect to that current task, and examined how these timing differences affected current task performance. We, however, concentrate on a common focal task performed over time (i.e., CABG), study *how* exposure to variety takes place with respect to past focal tasks—in conjunction with it (concurrently) or independently from it (non-concurrently)—and examine subsequent performance implications. Third, we make an important distinction between short-term and long-term learning dynamics. We introduce short-term exposure to variety as a factor that moderates how long-term exposure to variety (both concurrent and non-concurrent) affects individuals' focal productivity.

In the next section, we describe our setting before proceeding to motivate and develop our hypotheses.

2. Setting

Our setting is the cardiac unit of a private hospital in Europe, and our data set consists of 3,275 coronary artery bypass graft (CABG) surgeries that were conducted in the hospital over a period of seven years and three months. In addition to CABG surgeries, which we identify as the focal task (see Section 4 for a discussion), we also have information regarding all other types of cardiac surgeries that were conducted in the hospital during the same time period, which allows us to observe surgeons' exposure to other types of tasks (i.e., task variety).

Our setting is a very suitable context in which to investigate the effects of exposure to variety on individual learning and productivity, because surgeons perform CABG surgeries as well as a variety of other types of cardiac surgery and can therefore potentially transfer knowledge from one task to another. In addition, learning is an integral part of hospital operations (Tucker et al. 2007) and surgeons' practices (KC and Staats 2012). Furthermore, CABG surgeries are highly critical and complex yet quite common and frequent tasks for surgeons (Pisano et al. 2001, Clark and Huckman 2012), making them an ideal setting in which to study how exposure to task variety may influence learning and the resulting productivity of professional service workers. Finally, because our data set covers a span of more than seven years, we are able to examine both long-term and short-term effects of task variety on learning and subsequent focal productivity.

As mentioned, one major advantage of our setting is that the nature, type, and time of the tasks (surgeries) are not endogenously determined by the worker but are driven by outside factors (patients' needs). Consequently, a surgeon may perform a single CABG surgery if that is all the patient needs;

she may perform a single valve replacement if the patient requires only that, or she can perform multiple surgeries during the same operation (valve replacement and CABG) if the patient needs both. Likewise, surgeons' assignments to tasks and the similarity of their skills sets are other important considerations to avoid endogeneity; we investigate these in detail in the robustness checks section.

We identify CABG as the focal surgery in this setting and examine the impact of surgeons' exposure to other types of surgeries on subsequent focal task (CABG) productivity. Because surgeons perform other types of surgeries, too, both concurrently (e.g., performing valve replacement and CABG together) and non-concurrently (e.g., performing a single valve replacement) depending on medical requirements, we are able to observe knowledge transfer and learning between both concurrent and non-concurrent tasks.

3. Literature Review and Theory Development

3.1 Exposure to Variety, Learning, and Productivity

Previous research on the effect of task variety on individuals' performance has yielded mixed results. A significant body of literature argues that task variety enhances learning through successful transfer of knowledge between the focal and related activity (Tversky 1977 Zollo and Reuer 2010). In addition, task variety can increase workers' commitment and motivation (Hackman and Oldham 1976, Langer 1989), resulting in improved productivity. Schilling et al. (2003) showed that related variety can improve the learning rate of students playing different versions of a game. Staats and Gino (2012) showed that variety promotes workers' productivity in the long run.

On the other hand, researchers have also argued that task variety could be distracting and eventually detrimental to individual learning and productivity (Allport et al. 1994, Monsell 2003). Prior learning in one task may have a negative effect on performance of a related task due to 'negative transfer effect' (Gick and Holyoak 1987, Reed 1989, Zollo and Reuer 2010). Others have observed more nuanced relationships. For example, Narayanan et al. (2009) observed that variety had an inverted U-shaped relationship with individual productivity in an offshore software support services company, and KC and Staats (2012) found that focal subtask variety has an inverted U-shaped relationship with performance, whereas related subtask variety has a U-shaped relationship with the performance of cardiac surgeons.

3.2 Concurrent and Non-Concurrent Exposure to Variety

Our paper contributes to this debate by suggesting and testing a new mechanism by which task variety may influence individual learning and focal productivity. We reconcile the two views outlined above and suggest that exposure to variety independently (non-concurrently) or concurrently determines the direction and level of successful knowledge transfer from the related task to the focal task. That is, non-concurrent task variety could result in negative knowledge transfer and hence lead to lower productivity the next time the focal task is performed. On the other hand, concurrent variety could enhance knowledge transfer from related tasks to the focal task and hence improve individual productivity for

subsequent focal task.[1] For example, during an operation, a surgeon might perform only a valve replacement or might perform a valve replacement combined with a CABG. In the former operation, she will be exposed to variety (i.e., valve replacement) non-concurrently with the focal task (i.e., CABG), whereas in the latter case her exposure to variety will occur concurrently with the focal task.

3.3 Non-Concurrent Exposure to Variety

The basic mechanism through which past related tasks affect subsequent performance of a focal task is knowledge transfer. Negative transfer effects between tasks have been well established in the cognitive psychology literature (see Gick and Holyoak, 1987 for a review): prior learning and experience in one task may produce negative transfers and hence may be misleading in executing another task.

The most well-known mechanism for knowledge transfer is structure-mapping (Gentner 1983, Reed 1987, Ross 1987, Getner and Markman 1997): When faced with a task, individuals will compare it with past related tasks and retrieve past schema based on the similarities between the current task and past ones (Loewenstein et al. 1999). In explaining the mapping process, researchers have highlighted a critical distinction between surface similarity (i.e., similarity of context, goals, and processes) and structural similarity (i.e., deeper association in underlying causal structure). Even though the specific elements involved in tasks may resemble each other (surface similarity), the underlying causal structure and ordering of elements could be quite different (Novick 1988), and therefore, deriving causal inferences based on surface similarity can be misleading (Holyoak and Koh 1987).

Surface similarity has been found to have greater influence than structural similarity on schema retrieval from previous tasks (Holyoak and Koh 1987, Gentner 1989, Gentner and Calhoun 2010, Loewenstein et al. 1999). Indeed, individuals tend to focus on superficial (surface) details during the retrieval process, which is an important cognitive challenge of applying learning from one case to another (Loewenstein et al. 2003). Therefore, when individuals perform two tasks non-concurrently, they can be misled by the previous task's surface similarity, retrieve incorrect schema, and make inaccurate transfers if the tasks' surface similarity does not correlate well with their structural similarity (Reed 1987, Ross 1987, Ben-Zeev and Star 2001, Lee and Simon 2004). These incorrect transfers in turn lower the individual's productivity (Gick and Holyoak 1987).

Similarly, in our setting, while different types of cardiac operations are similar on the surface, there are important structural differences between them, which we argue would lead to a negative transfer. Specifically, in addition to CABG, our setting involves other types of operations, from tumor removal to heart transplant. Such operations share surface similarities with CABG, including context, aim, and

---

[1] Note that the question we examine is fundamentally different from that of Staats and Gino (2012), who examine the performance implications of "same-day different task" and "all prior days' different tasks". By contrast, we build on the observation that knowledge workers can perform various tasks either separately (i.e., non-concurrently) or together (i.e., concurrently) and highlight how this differential way of getting exposure to different tasks influences subsequent productivity on a common focal task.

process, but they are carried out in distinct steps with significant structural differences from CABG (Reznick and MacRae 2006)[2]. During a CABG operation, a vein from elsewhere in the patient's body is stitched to the aorta and a coronary artery, whereas during a tumor removal, for instance, the tumor is removed from the heart or the valve along with the tissue around it.

Hence despite the surface similarities, we argue that different cardiac surgery types do not have enough structural similarity to allow for accurate causal inferences from one another (Holyoak and Koh 1987) when they are performed at different times (non-concurrently). Specifically, non-concurrent variety will not enable the surgeons to directly compare (Loewenstein et al. 1999) and successfully comprehend the alignable differences (i.e., differences that are connected to the common context of the two tasks) (Gentner and Markman 1997) between the two operation types. When variety is non-concurrent (i.e., the other type of task is performed at an earlier time), surgeons are likely to retrieve schema that are not exactly appropriate for the present task, leading to inaccurate transfer and therefore reduced productivity. Such inaccuracy is due to the dominance of surface similarity in schema retrieval from long-term memory (Gentner and Calhoun 2010, Loewenstein et al. 1999).

In addition to negative transfer effects, the development of implicit memory (i.e., priming) due to non-concurrent variety can impair decision making in a subsequent focal task (Allport and Wylie 1999). Medical researchers have found that surgery operations require 25 percent technical and 75 percent decision-making skills (Spencer 1978, Grierson et al. 2011) and that doctors and medical staff do respond unconsciously based on past stimuli (Loewenstein and Lerner 2002, Bargh and Williams 2007). These findings suggest that priming from related variety can lead surgeons to take suboptimal decisions and actions in a subsequent focal surgery (i.e., CABG) and thereby decrease their focal productivity. For these reasons, we predict that:

Hypothesis 1: *Non-concurrent exposure to task variety has a negative impact on subsequent focal task productivity.*

3.4 Concurrent Exposure to Variety

We next examine the productivity implications of performing other tasks in conjunction with the focal task. When another task is performed concurrently with the focal one, surface similarities between tasks will not mask their structural differences, because individuals will be able to compare and comprehend alignable differences between different tasks (Gentner and Markman 1997). This comparison entails a mapping process that highlights both the differences and similarities between the two tasks (Holyoak and Thagard 1995, Medin et al. 1993, Gentner and Markman 1997) and results in the abstraction of schema that are useful for the focal task (Lassaline and Murphy 1998, Markman and Gentner 1993).

Individuals also better comprehend, learn about, and identify the intricacies of the focal task as a result of the greater implicit learning (Reber 1989, Wulf and Schmidt 1997) and discrimination

---

[2] We also investigate potential effects of the differences between non-CABG operations in the robustness checks section.

(Anderson 1982) that come from concurrent variety. Indeed, concurrent variety provides a new, varied context for the focal task which promotes implicit learning. As a result, individuals develop critical but highly complex and abstract knowledge about the focal task and its associations (Maskarinec and Thompson 1976). In addition, concurrent variety facilitates the discrimination process necessary for cognitive learning (Anderson 1982) by producing multiple variants of the conditions of the same action. Performing two tasks concurrently activates cognitive processes (Koechlin et al. 1999) which are linked with the ability to integrate and combine subtasks and ideas (Koechlin and Hyafil 2007). Individuals can then identify the most appropriate course of action by learning about and comparing differences between the focal task and the concurrent task (Anderson 2013).

For example, the common reference book for cardiac surgery—Kirklin/Barrat-Boyes (Kouchoukos et al. 2012, p. 529)—identifies two primary considerations when CABG is accompanied by mitral valve repair/replacement:[3] *"1. Planning the operation to limit the duration of CPD and global myocardial ischemia (aortic clamping), 2. Reducing the need to tilt the heart up after the valve is inserted or repaired."* These specifications suggest that concurrent variety can facilitate processes of implicit learning and discrimination by giving the focal CABG operation a different context and differing conditions in the form of additional coordination and planning considerations and new constraints. Considering all the above arguments, we predict that:

Hypothesis 2: *Concurrent exposure to task variety has a positive impact on subsequent focal task productivity.*

3.5 The Moderating Role of Short-Term Non-Concurrent Variety

We next ask how recent (short-term) non-concurrent variety interacts with long-term non-concurrent variety. We argue that the negative effects of non-concurrent variety will be aggravated by recent non-concurrent exposure to variety. The negative knowledge-transfer effect (Gick and Holyoak 1987) of non-concurrent variety will be further amplified by recent non-concurrent exposure to variety. This is because individuals tend to retrieve schemas that they have used recently, even when more plausible and reasonable alternatives exist (Reder 1982). So, a recent non-concurrent exposure to variety will lead surgeons to adopt its schema, which may not be appropriate for the current task.

In addition, as discussed in Section 3.3, non-concurrent exposure to variety may decrease surgeons' productivity due to their response to similar stimuli, which may prove detrimental over time. Researchers have shown that priming is more likely to happen when exposure to the related task is more recent (Lerner et al. 2004, Bargh and Williams 2006). Compared with more distant experiences, the carryover effects of recent experiences are more likely to cause automatic responses to subsequent

---

[3] We use valve operation as the illustrative example because previous literature has highlighted the common occurrence of CABG and valve operations (Ch'ng et al. 2015, Kimiyoshi et al. 2009, Saxena et al. 2013). Indeed, valve repair/replacement is the most frequent concurrent operation with a CABG in our sample.

similar experiences (Bargh and Williams 2006), since individuals tend to respond unconsciously with their most recent relevant behaviors (Bargh et al. 2012). We therefore expect that:

Hypothesis 3: *Recent non-concurrent exposure to variety amplifies the negative effect of non-concurrent exposure to variety on subsequent focal task productivity.*

3.6 The Moderating Role of Short-Term Concurrent Variety

Finally, we consider the moderating effect of recent (short-term) concurrent exposure to variety on the relationship between concurrent exposure to task variety and productivity. We expect that the positive knowledge transfer from all past concurrent exposures to variety will be higher after a recent exposure to concurrent variety.

As discussed in Section 3.4, an important way that concurrent variety improves subsequent focal task productivity is implicit learning. Because this process is essentially about identifying and recalling associations of the focal task (Maskarinec and Thompson 1976), which implicitly involves time, exposure to recent concurrent variety will further improve implicit learning.

In addition, individuals tend to forget (Shtub et al. 1993, Egelman et al. 2016), which can have a significant impact on their productivity in procedural cognitive tasks (Bailey 1989, Nembhard and Uzumeri 2000) such as cardiac operations. This forgetting effect becomes less rapid as the complexity of the performed task increases (Lance et al. 1998, Nempbhard 2000) and more rapid as the time interval between two consecutive tasks increases (Globerson et al. 1989, Bailey 1989), even in a procedural cognitive task (Nembhard and Uzumeri 2000) such as a cardiac operation. Researchers have also shown that surgeons' learned skills (both technical and cognitive) tend to deteriorate over time (Ramdas et al. 2017) and have highlighted the need for periodic remediation of any necessary skills (Kahol et al. 2010). Moreover, Kahol et al. (2010) have suggested that surgeons can improve their retention of skills by performing multiple surgeries concurrently. Hence we believe that a recent concurrent exposure will significantly reduce the forgetting effect for a surgeon. With reduced forgetting, long-term learning from exposure to variety will increase. Therefore, we expect that:

Hypothesis 4: *Recent concurrent exposure to variety amplifies the positive effect of concurrent exposure to variety on subsequent focal task productivity.*

4. Data and Variables

The organization that we use for our study is the cardiac unit of a private hospital in Europe that is the property of an American non-profit organization. The hospital admits more than 2,000 patients annually and performs around 850 cardiac operations each year. We test our hypotheses using an archival data set of all 3,275 CABG operations performed in the hospital during the period from 01/01/2004 to 31/03/2011. After removing two operations with missing data, we are left with 3,723 operations for our study.

Each surgery team consists of one lead surgeon and zero to four assistant surgeons. Our sample includes 44 surgeons, 19 of whom started working after the beginning of our dataset, and 13 of whom do not appear during the last year of our dataset. Apart from the surgeons, each team also typically has one anesthesiologist, one perfusionist, and zero to three scrub nurses. We provide further information on surgeons in the appendix. Like other studies that examine the effect of experience on performance in surgical settings (see, for example, KC and Staats 2012), our study concentrates on surgeons (lead and assistant surgeons, 44 in total) and their exposure to various tasks. There are two reasons for this. First, surgeons perform different sets of steps and activities during different kinds of surgery, which gives them many opportunities to learn. The rest of the team members (e.g., the anaesthesiologist preparing the patient, nurses providing the equipment, etc.) perform tasks that are more trivial and very similar across different kinds of surgery. Since our primary agenda in this paper is to identify the effect of task variety on learning and productivity, we focus on surgeons, who perform somewhat different activities and tasks in different kinds of surgery. Second, interviews with the medical staff at the hospital confirmed our intuition that we should consider only surgeons in studying the effect of task variety on productivity in this setting.

Our data set contains information about the type of operation performed, the members of the surgical team, and the operation's duration, including exact start and end times. Our sample also includes information regarding the patient's age, sex, and condition before the operation. Specifically, the hospital labels each patient's case either "severe", "medium" or "mild". Finally, our data set includes limited information about in-hospital mortality for patients who have had an operation in the hospital.

To test our hypotheses, we use CABG as the focal surgery type and examine the impact of exposure to other cardiac surgery types (i.e., concurrently vs. non-concurrently) on the duration of subsequent CABG-only surgeries. We use CABG as our focal task because it is the most common cardiac surgery type (Clark and Huckman 2012), and indeed it appears more than any other surgery type in our data set. In addition, since our goal is to examine the different effects of concurrent and non-concurrent task variety on subsequent productivity in the focal task, we need a task which could be performed both concurrently with another task and on its own. Finally, CABG surgeries have received significant attention in the recent operations management literature (Pisano et al. 2001, Huckman 2003, Huckman and Pisano 2006, KC and Staats 2012), which could help us consider our results' validity and generalizability. Consequently, CABG is an ideal choice of focal task for our study.

A surgeon can perform multiple types of surgery on the same patient during the same operation, as required by medical conditions. For this analysis, in addition to the focal task (i.e., CABG), we have information regarding other types of cardiac surgeries performed during the same time interval in the hospital. Our data set includes 1,324 valve repair/replacements, 86 congenital surgeries, 70 heart failure procedures, 20 tumour removals, 185 routine cardiac surgeries, and 78 other normal surgeries (all other surgeries that do not fit into any of the previous categories). In addition, our data set includes 951 complex operations in which a CABG surgery and one of the other types of surgeries were performed

concurrently. The data set also includes 170 very complex operations in which a CABG and two of the other types of surgeries were performed concurrently. Because all our hypotheses address focal task productivity, we use 3,273 CABG-only surgeries as our observations to test the hypotheses. However, when calculating our independent variables, we make use of all 6,171 surgeries (CABG only, other type only, and concurrent surgeries; it is worth noting that apart from the 951 complex operations which include a CABG, 14 additional operations include two surgeries other than CABG performed concurrently); details appear in Section 4.4.

Finally, we also conducted a limited number of interviews with several hospital staff members. These interviews enabled us to better comprehend how CABG and other surgeries are performed and helped us understand hospital policies and management practices.

4.1 Variables

Dependent Variable. We use duration of CABG operations as our dependent variable. Extensive research in psychology has shown that decrease of the completion time of a performed task is an indicator of learning and increased productivity (Thurstone 1919, Graham and Gagne 1940). In examining performance implications of experience and learning-related issues, operation completion time is a commonly employed dependent variable. For example, in their investigation of the effect of team familiarity, organizational experience, and role experience on team productivity, Reagans et al. (2005) employed surgery duration as the dependent variable. Similarly, other learning-related studies such as Pisano et al. (2001) and Edmondson et al. (2003) used procedure completion time for cardiac surgery as their dependent variable. Similarly, we contend that lower completion time reflects learning and increased productivity for the surgeons and so use it as our dependent variable. In addition, consistent with the previous literature which found that shorter completion times decrease the probability of post-surgical infections for cardiac surgeries (Gaynes et al. 2001, Gibbons et al. 2011), in our semiformal interviews, staff members in our hospital also confirmed that lower completion times usually reflect better clinical outcomes.

While our data set also included in-hospital mortalities, unlike in large-scale multi-hospital studies, death events were quite infrequent in our setting of only one unit of a single hospital. Consequently, like Pisano et al. (2001), we were not able to detect any significant variables that explain variation in mortality rates other than the clinical condition of the patient (i.e., severe, medium, mild). In fact, all mortalities came from the severe group. Therefore, in line with our theoretical development, we have decided to keep the scope of our study on surgical productivity implications of concurrent and non-concurrent task variety and leave it to future work to explore quality (e.g., mortality) implications of task variety.

4.2 Independent Variables

Non-Concurrent Variety. This variable captures all prior days' noncurrent exposure to variety for the surgeons in a team. For each surgeon in a team, we first count the total number of times since the

beginning of our dataset that she has performed a surgery that does not include CABG (that is, other type only), up to the current CABG operation. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Concurrent Variety. This variable captures all prior days' concurrent exposure to variety for the surgeons in a team. For each surgeon, we first count the total number of concurrent operations she has performed, that is, an operation which involves a CABG surgery and one of the other types of surgeries up to the current (focal) CABG surgery since the beginning of our dataset. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Non-Concurrent Variety x Recent Non-Concurrent Variety. We first calculate Recent Non-Concurrent Variety. To capture Recent Non-Concurrent Variety, we calculate for each surgeon the number of surgeries that did not include a CABG operation (that is, other type only) that she has performed during the week before the current CABG surgery. We then multiply this new variable by the Non-Concurrent Variety variable to create the interaction term.

Given our setting, we use one week to capture a short time period (i.e., recent). Clearly, what constitutes a short time period may differ depending on the operational context. Given operational dynamics in our context and the fact that CABG operations last around three to six hours, we think that one week is an appropriate choice. (See also Staats and Gino 2012 p. 143 for a similar discussion, which suggests that a week may be an appropriate choice for characterizing a short time period with tasks that last five hours). Consequently, in our study, any exposure to variety that takes place within the week prior to the current operation is considered recent. We also performed robustness checks for our choice of one week by considering slightly longer and shorter time periods, and our results remained the same.

Concurrent Variety x Recent Concurrent Variety. We first calculate Recent Concurrent Variety. To capture Recent Concurrent Variety, we calculate for each surgeon the number of concurrent operations (a CABG and another type concurrently) she has performed during the week before the current CABG surgery. We then multiply this new variable with the variable Concurrent Variety to create the interaction term.

4.3 Control Variables

Focal Experience. For each surgeon, we calculate the number of CABG surgeries (either CABG-only or combined with another surgery type) she has conducted up to the current CABG surgery (excluding the current one) since the beginning of our dataset. This way, we capture each surgeon's total experience with the focal surgery type. Please also see Table 1 for an illustration of how we calculate this variable for an individual surgeon.

Team Size. We control for the size of the surgical team (counting all team members). Larger teams might have more access to experience and resources (Reagans et al. 2005), but smaller work-group size is associated with increased team productivity (Gladstein, 1984), since larger teams sometimes face coordination challenges that decrease their productivity (Hackman 2002).

Time Fixed Effect. To control for potential environmental changes in our setting (such as hospital policy or technological advances) and also organizational experience that may influence surgery durations, we include dummy variables indicating the year, month, and day of the week that the operation took place.

Individual Average Experience. We control for the average experience—measured in number of operations—of the team members other than the lead surgeon. For each team member (excluding the lead surgeon), we calculate the number of times she appears in any operation prior to the current one (not including the current one) since the beginning of our dataset. We then take the sum and divide by the number of team members (excluding the lead surgeon).

Indicators for Severity of the Case. As mentioned, the hospital labels each patient's case mild, medium, or severe. The "medium" category appears most often in our sample, so we include two dummy variables in all models: "Severe" and "mild" are both equal to one if the hospital has labelled the patient as such and zero otherwise. We expect "severe" cases to generally last longer than the other two categories, as indicated by table 1 in the Appendix.

Age. For every patient we control for age since operations may last longer for older patients.

Male. Finally, we also control for the gender of each patient by including a dummy variable equal to one if the patient is male and zero otherwise.

4.4 Calculation of Key Independent and Control Variables

Table 1 shows how we calculate Focal Experience, Concurrent Variety, and Non-Concurrent Variety for each surgeon in our sample. As mentioned, one operation may include more than one surgery type. That is, while the most typical case is to perform only one type of surgery (e.g., only a CABG or only a valve replacement), a good number of operations include more than one surgery type (e.g., a CABG plus a valve replacement), depending on the patient's medical requirements. It is also worth noting that when we calculate Non-Concurrent and Concurrent Variety, we remove the corresponding recent varieties to avoid confounding our variables[4].

At time t_1, the surgeon performs an operation that includes a CABG and a valve repair. Then in our first observation at time t_2 (our observations are the operations that include only a CABG surgery), her score for Focal Experience will be equal to 1, since she has conducted one CABG surgery prior to t_2. Similarly, her score for Concurrent Variety will be equal to 1, because she has conducted a valve repair and CABG concurrently prior to t_2. Finally, her score for Non-Concurrent Variety will be equal to 0, since she has not yet conducted any other type of surgery individually (non-concurrently). Then at time t_3 she performs an operation that includes only a valve repair. In our next observation for our study at time t_4, her score for Focal Experience will become 2, because she has conducted two CABG surgeries

---

[4] We also repeat our analysis without removing corresponding recent variety and get very similar results.

(at times t_1 and t_2) prior to t_4. Her score for Concurrent Variety will remain 1, and her score for Non-Concurrent Variety will now become 1 because she has conducted an individual valve repair surgery on a patient prior to t_4 (i.e., at time t_3).

Table 1. Surgeries performed by an individual surgeon over time (t=t_1, t_2, t_3, t_4)

Time	Patient	Performed Surgery		Independent Variables			Dependent Variable
		CABG	Valve Repair	Focal Experience	Non-Concurrent Variety	Concurrent Variety	
t_1	Patient 1	√	√	n/a	n/a	n/a	
t_2	Patient 2	√		1	0	1	Observation 1
t_3	Patient 3		√	n/a	n/a	n/a	
t_4	Patient 4	√		2	1	1	Observation 2

5. Empirical Approach and Results

We test our hypotheses using a fixed-effects panel regression with AR(1) disturbance based on Baltagi and Wu (1999). That is, we examine the effect of concurrent and non-concurrent variety for lead surgeon i on the duration of CABG operation j. Our panel data structure, in which surgeons perform operations over time, poses several challenges. First, there may be unobserved heterogeneity between surgeons in areas such as ability, education, or experience, and this heterogeneity may bias our results. Second, there is potential serial correlation between operations that were performed by the same lead surgeon close in time. In fact, when we use the Durbin-Watson test, we find that autocorrelation in our sample is likely to be first-order (i.e., AR(1)). Third, we have an unbalanced panel structure with unequally spaced observations over time. Our fixed-effects regression with AR(1) disturbance based on Baltagi and Wu (1999) (i.e., xtregar, fe command in STATA) explicitly takes into account all these issues and eliminates all unobserved time-invariant individual effects of lead surgeons.

Also, our analyses of the distribution of dependent and continuous independent variables (*ladder* and *gladder* commands in Stata) reveals skewed distributions. Consequently, and in line with our theoretical arguments and previous studies of the use of conventional learning-curve models to examine completion times (Argote 1999, Reagans et al 2005), we take the logarithm of all these variables (apart from age). We also confirm the normality of residuals and check for heteroscedasticity by using the Breusch-Pagan test (1979), which does not reject the null hypothesis, thereby confirming that heteroscedasticity does not pose a threat to our analyses.

Our model is the following:

$ln(duration_{ij}) = \beta_0 + \beta_1 \, ln(Non\text{-}Concurrent \; Variety_{ij})$

$$\beta_2 \, ln(Concurrent \; Variety_{ij}) +$$

$$\beta_3 \, ln(Recent \; Non\text{-}Concurrent \; Variety_{ij}) \; x \; ln(Non\text{-}Concurrent \; Variety_{ij}) +$$

$$\beta_4 \, ln(Recent \; Concurrent \; Variety_{ij}) \; x \; ln(Concurrent \; Variety_{ij}) +$$

$$\beta_5 \, ln(Recent \; Non\text{-}Concurrent \; Variety_{ij}) +$$

$$\beta_6 \, ln(Recent \; Concurrent \; Variety_{ij}) +$$

$$\beta_7 \, ln(Focal \; Experience_{ij}) +$$

$$\beta_8 \, ln(Team \; Size_{ij}) +$$

$$\beta_9 \, ln(Individual \; Average \; Experience_{ij}) +$$

$$\beta_{10} \, (Severe_{ij}) +$$

$$\beta_{11} \, (Mild_{ij}) +$$

$$\beta_{12} \, (Age_{ij}) +$$

$$\beta_{13} \, (Male_{ij}) +$$

$$\alpha_i +$$

$$t_j +$$

$$u_{ij} \, ,$$

$$\text{where} \;\; u_{ij} = \rho u_{ij\text{-}1} + e_{ij}$$

In the above model, α_i represents unobserved lead surgeon characteristics such as education and background. Our fixed effects estimation eliminates α_i by de-meaning the variables using the within transformation. Also, in the model, t_j represents time effect, indicating the year, month, and day of the week that the operation takes place. In addition, u_{ij} is the serially correlated error term, with $|\rho| < 1$ being the first-order autocorrelation coefficient, and e_{ij} is independent and identically distributed with zero mean and constant variance.

Table 2 shows descriptive statistics and correlations among the dependent, independent, and control variables for the lead surgeons. Table 3 shows the results for all our hypotheses for the lead surgeons[5]. Due to the AR(1) covariance structure we employ, the number of observations in the table is equal to 3,261. In Model 3.1 we include only our control variables. As expected, *focal experience* and *average individual experience* have negative and significant coefficients, suggesting that they reduce completion times, whereas team size increases duration. In addition, compared with the baseline group of medium, *severe* operations take longer and *mild* operations are shorter in duration.

In Model 3.2 (Table 3), we add our first variable of interest: *Non-Concurrent Variety*. The adjusted R^2 is increased by 3.45%, and an F-test shows that Model 3.2 is superior to Model 3.1 ($p < 0.01$). *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$), providing support for our first hypothesis. In Model 3.3 (Table 3) we add *Concurrent Variety*. The adjusted R^2 is further increased by 8.00%, and an F-test shows that Model 3.3 is superior to Model 3.2 ($p < 0.01$). *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$), supporting Hypothesis 2.

---

[5] Please see Section 2 in the appendix for the analysis with all surgeons.

In Model 3.4 (Table 3) we add the interaction terms *Non-Concurrent Variety x Recent Non-Concurrent Variety* and the variable *Recent Non-Concurrent Variety*. The adjusted R^2 is further increased by 1.85%, and an F-test indicates that Model 3.4 is superior to Model 3.3 ($p < 0.05$). We also see that *Non-Concurrent Variety x Recent Non-Concurrent Variety* is significant ($p < 0.01$) and positive, providing support for Hypothesis 3. Finally, in Model 3.5 (Table 3) we include the interaction term *Concurrent Variety x Recent Concurrent Variety* and the variable *Recent Concurrent Variety*. The adjusted R^2 is further increased by 1.82%, and an F-test indicates that Model 3.5 is superior to Model 3.4 ($p < 0.05$). *Concurrent Variety x Recent Concurrent Variety* is significant ($p < 0.05$) and negative, providing support for Hypothesis 4[6].

According to Model 3.3 (Table 3), an increase of 20% in *Non-Concurrent Variety* multiplies expected duration by $e^{0.312*\ln(1.2)} = 1.0585$, that is, duration is increased by 5.85%. The average duration in our sample is 297.77 minutes. Hence this increase translates into 5.85%*297.77 = 17.420 minutes. Similarly a 20% increase of *Concurrent Variety* decreases duration by 6.03% (17.956 minutes). This 20% increase corresponds to,[7] on average, around 32 (=159.91*0.2) non-concurrent and 27 (=136.88*0.2) additional operations for each lead surgeon. In our sample, on average, each lead surgeon performs 0.74 non-concurrent operations and 0.57 concurrent operations per week. Hence, based on existing surgery frequencies in our sample, it takes a lead surgeon roughly 32/0.74=43 weeks and 27/0.57=47 weeks to achieve a 20% increase for non-concurrent and concurrent variety respectively. These numbers imply that lead surgeons can indeed achieve meaningful productivity improvements within less than a year.

One limitation of our fixed-effects AR(1) regression based on Baltagi and Wu (1999) is that it is not possible to report robust standard errors. The Breusch-Pagan test (1979) reveals no heteroscedasticity, but, nevertheless, we additionally run fixed-effects regression with robust standard errors (*xtreg fe* with *robust* option) and test all our hypotheses. Tables A8 and A9 in the appendix show the results for our hypotheses using this alternative model. The results for all our hypotheses remain the same in terms of significance and support and are close in terms of coefficients.

6. Robustness Checks

We perform several additional analyses to examine the robustness of our results and to rule out potential alternative explanations.

6.1 Data and Variables' Operationalization

---

[6] While the interaction effects for Hypotheses 3 and 4 are statistically significant in all models, post hoc plots (Figures A3 and A4 in the appendix) reveal that the effect of the moderation is quite modest in terms of economic significance (also note the coefficients of the interaction terms in our models).

[7] According to Table A3 of the appendix, *Non-concurrent Variety* and *Concurrent Variety* have average values of 159.51 and 136.88 respectively.

As in any empirical study, our data set is limited. A potential concern is that we have no data for surgeries or surgeons prior to the beginning of our data set, which may influence our results. To address this, we repeat our analysis after excluding different time intervals from the beginning of our data set. That is, we remove the first 12 months and 24 months, and we calculate our dependent variable and all independent variables using the remaining data (e.g., when we remove 24 months, we calculate surgery durations using only those surgeries that took place between months 24 and 87, and when calculating concurrent exposure to variety, for instance, we similarly use all operations between months 24 to 87). We then repeat our analysis. The results for all our hypotheses remain the same quantitatively (i.e., the sign of the coefficients for our variables of interest remain the same and significant either at 1% or 5% level). Finally, Lapre and Tsikriktsis (2006) suggest that studies examining a learning effect a long time after the beginning of the learning curve can eliminate potential bias by using a log-linear model. Consequently, we repeat our analysis for H1 and H2 by using a log-linear model and find full support for all our hypotheses. We therefore believe that missing experience (i.e., missing data prior to the beginning of our data set) does not pose a threat to our main findings.

We also test the sensitivity of our results by repeating our analysis after removing the first 12 months of our dataset only for the dependent variable (about 14% of our initial observations). That is, while we include our entire dataset when calculating the independent variables, in this case the set of observations (and hence the dependent variable) start after month 12. This way, we investigate the robustness of our results when there is missing data for our main independent variables at the beginning. While the magnitude of effects change as expected, our primary findings remain quantitatively the same.

We also replace Average Individual Experience of the rest of the team members with a variable called Average Individual Direct Experience, which captures the number of times each non-surgeon team member has conducted the focal type of operation (CABG) since the beginning of our data set. Our results remain the same.

One may also be concerned that our observed effects may only work on single focal operations (CABG-only) and may not hold for non-focused operations that involve more than just CABG. In order to examine this, we repeat our analysis after changing the way we define our variables. Specifically, we define as focal experience all CABG and valve repair/replacement surgeries (the second most frequent operation type in our sample) and calculate *Focal Experience, Non-Concurrent Variety,* and *Concurrent Variety* accordingly. Specifically, we use all operations that include a CABG or a valve repair/replacement when calculating *Focal Experience*, all operations that include neither CABG nor valve repair/replacement when calculating *Non-Concurrent Variety,* and all operations that include both CABG and valve repair/replacement and another type when calculating *Concurrent Variety.* We repeat our analysis and find support for all our hypotheses. This confirms that the effects of non-concurrent and concurrent exposure to variety hold in operations that include more than a CABG.

Moreover, we repeat our analysis using all operations (the whole sample) as observations and not just the CABG operations. This is because if our results pertaining Non-Concurrent Variety change

significantly and we observe an exactly opposite effect, then this could imply very different managerial implications for organizing Non-Concurrent Variety. In this new analysis on the whole sample, depending on the currently performed operation, we define and calculate our variables of interest accordingly. If, for example, the observation is a single valve operation, then we define all valve operations in the past as focal experience, all non-valve operations in the past as non-concurrent variety, and all operations that include a valve operation and a second type as concurrent variety. Then we perform a single analysis in our whole dataset and get support for our hypotheses. Specifically, our results remain similar to those we obtained with the CABG-only sample, though with smaller effect sizes for both Non-concurrent variety and Concurrent variety. Hence, we believe that our results do not depend on the specific choice of focal operation.

Furthermore, one may argue that some non-CABG operations are more similar to CABG than others and that this similarity might drive the results regarding the effects of concurrent and non-concurrent variety. In order to address this, we group non-CABG operations into two broad categories: i) those that are more related to CABG and ii) those that are less related to CABG, based on previous medical and management literature. In the former category, we include only valve operations, and in the latter category, we include the rest of the non-CABG operations. Medical literature has pointed out the overlap between CABG and valve operations in terms of the surgery process (Kouchoukos et al. 2012, Ch'ng et al. 2015), an assessment which was also echoed by studies in management which highlighted the similar set of moves during both surgeries (Pisano et al. 2001). On the other hand, other types of surgeries such as tumour removal or congenital surgery have been suggested to be quite different from CABG in terms of the steps and processes involved (Kouchoukos et al. 2012). For our analysis, we break down non-concurrent variety into more-related non-concurrent variety and less-related non-concurrent variety, and similarly concurrent variety into more-related concurrent variety and less-related concurrent variety. Next, we repeat our analysis for the lead surgeons using only more-related non-concurrent variety and more-related concurrent variety in one model and only less-related non-concurrent variety and less-related concurrent variety in another model. We report corresponding results in Table A10 and A11 in the appendix. Our results are consistent in both models, and they remain similar to the original results (i.e., without breaking down non-CABG operations), providing further support for our hypotheses on the effects of concurrent and non-concurrent variety. That is, regardless of being more related or less related to CABG, concurrent non-CABG operations are associated with decreased duration, whereas non-concurrent non-CABG operations are associated with increased duration.

Finally, while our variables *focal experience, concurrent variety, and non-concurrent variety* capture the experience of a surgeon during the time interval of our data set, a potential concern could be that we do not specifically account for potential tenure/experience differences of surgeons and associated unobserved factors (e.g., familiarity with the hospital environment), as we do not have detailed personnel data. In order to address this, we conduct additional analyses in which we only

consider the subsample of operations which only includes the lead surgeons who have been at the hospital from the beginning of our data set. This way, we are able to minimize and control for any potential experience-based differences across surgeons, as these surgeons all have been at the hospital for more than seven years, suggesting they are all highly familiar with the hospital environment. Our results with this more conservative analysis using the subsample provide full support for all our hypotheses (Table 4).

6.2 Surgery Assignments

Considering our study, there may be two types of potential selection issues that may bias our results. The first one is associated with patients' selection of specific doctors, and the second one is linked to doctors' selection or avoidance of certain patients or surgery types. One major advantage of our setting of the private European hospital is that patients by and large are referred to the hospital and not to a specific surgeon in the hospital. The hospital then assigns patients to surgeons. This mechanism in which the hospital plays a major role in assignment considerably alleviates patient-driven or surgeon-driven selection concerns as patient/surgeon assignments are plausibly random.

However, this mechanism may then arguably give rise to another potential selection concern in that the hospital may assign certain cases to certain doctors (e.g., more severe cases to more skilled, expert surgeons or easier cases to less experienced surgeons). If this is the case, then it could bias our results. This is unlikely to be the case in our setting, because all surgeons at this highly reputed private hospital are expected to be highly skilled and experienced and are expected to readily perform all types of surgeries. Our interviews with hospital administration and staff confirmed this. We also examine this concern empirically. Specifically, we first investigate the distribution of severe cases among surgeons and do not observe any patterns. Second, we conduct a chi-square test to ensure that the mild, medium, and severe cases are evenly spread across lead surgeons. The results (Table A5 in Appendix) indicate that there is no difference across lead surgeons in terms of severity of assignments. Next, we repeat our analysis after dropping the most critical cases from our sample, which includes patients that died in the hospital after the operation, and find the same results quantitatively. We also investigate the spread of deaths among surgeons and examine any potential correlations of these deaths with our key variables. We find that these in-hospital deaths are spread evenly across surgeons and show no correlation with our key variables other than the "severe" patient condition. These analyses suggest that there is no systematic difference across surgeons in terms of severity of the cases being assigned to them.

A severe case can be more critical than another severe case, and, similarly, one mild case might be easier than another. Because the hospital does not make these kinds of distinctions within severe cases and within mild cases, we further examine this issue: We repeat our analysis after excluding the severe cases (using only the mild and medium ones). Then we repeat our analysis after excluding only the mild cases from our initial sample (using only the medium and severe ones). Finally, we also repeat our analysis using only the medium cases. In all three cases, while the magnitude of effects changed

considerably as expected, our results remain the same (please refer to Tables A12-A17 in the appendix for corresponding results).

Finally, we further investigate the relationship between in-hospital mortality and duration. Specifically, we create two groups of observations: The first group includes all the operations that resulted in the death of the patient after the completion of the operation (i.e., post-operation in-hospital mortality; this is the only post-operation mortality information we have available to us). Also please note that all these operations were characterized as severe. The second group includes all the severe operations that did not result in the death of the patient. Please note that in the former group, we do not include the operations that resulted in the death of the patient during the operation, since surgery duration would be biased in these cases. We conduct a t-test to compare the average duration of these two groups. Our results indicate that there is no significant difference in the duration of these two groups (p-value= 0.16). Next, we create a more granular matched sample using the Coarsened Exact Matching algorithm described by Iacus et al. (2011) using the Stata command cem. This enables us to compare the duration of the operations in a more closely matched sample. Specifically, we focus only on the severe cases, exclude the cases in which the patient died during the operation, and use a dummy variable equal to 1 if the operation resulted in the death of patient after the operation and 0 otherwise as our treatment variable. We also use the covariates of age, gender, lead surgeon, and surgery type for our matching process. Our results indicate that 79% of our observations are matched and the L1 statistic—a widely accepted measure of global imbalance comparison (Iacus et al. 2011)—is reduced by 19.6% (from 0.449 to 0.361), indicating a significant decrease of imbalance after our matching process. The t-test after the matching process indicates no significant difference in duration between severe operations with mortality and severe operations with no mortality (p-value > 0.10). While this analysis is helpful, we should acknowledge that its extent is limited due to our limited data on patient characteristics and all mortalities observed in the severe group.

6.3 Instrumental Variables Approach

As discussed in the previous section, we believe that our particular empirical setting (where surgeon/patient assignment is plausibly random) and additional robustness checks address most selection concerns. In addition, our fixed-effects regressions accounts for all observed and unobserved time-invariant heterogeneity across lead surgeons through a de-meaning process. We also include a set of important control variables, such as the lead surgeon's focal (CABG) experience, other team members' average experience, team size, time effects, and indicators of the severity of the case, as well as patient age and gender, which may affect the duration of an operation. These, we believe, address most plausible endogeneity concerns with regards to our findings about concurrent and non-concurrent variety on surgery duration.

However, if there are unobserved time-varying individual factors that affect both lead surgeons' task variety and their surgery duration at the same time, then this may be a concern. In order to address

such potential concern, we conduct an additional instrumental variables analysis with 2SLS specification. Specifically, we identify and use two plausible instruments for surgeons' concurrent and non-concurrent variety based on surgeons' vacations in our context. We define vacations if the lead surgeon does not appear in our sample for one week or more (we also consider shorter and longer durations and obtained similar results) and then reappears (which means that she is still working at the hospital). Our surgeons work solely for our client hospital, so their absence in our sample does not indicate that they may be working in another hospital or in private practice.

The first instrument we use is *Others' vacation.*[8] For the focal lead surgeon i and operation j, this is calculated as the total number of vacation days of all other lead surgeons in the hospital from the time lead surgeon j joined the hospital until the current operation j.[9] The idea is that the higher the number of days other surgeons are on vacation, the greater the expected number of concurrent and non-concurrent tasks performed by the focal surgeon (e.g., in others' absence, there is plausible increase in operations performed by the focal surgeon as absent surgeons' operations will be distributed among present ones).[10] We can empirically verify this by checking the first-stage regression. However, other surgeons' previous days on vacation should not have any direct influence on the focal surgeon's current surgery duration, except through increasing her concurrent and non-concurrent task variety. That is, we expect this instrument to satisfy the relevance and exclusion conditions and have a significant positive effect on concurrent and non-concurrent task variety in the first stage.

The second instrument we use is the focal surgeon's vacation days in the past year, excluding the most recent 3 months. For ease of reference, we call this variable *Days of vacation*. For the focal lead surgeon i and operation j, this is calculated as the total number of vacation days of the focal surgeon i during one year[11] before the operation j, except the most recent 3 months. We expect that the higher the surgeon's number of days spent on holiday in the past year, the lower her concurrent and non-concurrent task variety, as the surgeon cannot perform operations during these periods. Also, in order to eliminate any possible direct effect of a surgeon's holiday on her surgery duration (e.g., warming up, adjustment periods), we exclude the most recent 3 months in our calculation of the instrument (which is the mean inter-holiday time of surgeons in our sample).[12] That is, while this instrument helps explain variation in a surgeon's concurrent and non-concurrent variety, it should not have any direct influence on her current surgery duration, therefore satisfying the exclusion condition.

---

[8] We include summary statistics for our instruments in the appendix.

[9] We also repeat our analysis by calculating *Others' vacation* using the beginning of our dataset for all lead surgeons and get the same results.

[10] Our hospital does not hire temporary surgeons, so in planning the absence of a surgeon, her anticipated operations are distributed among her colleagues.

[11] Our choice of one year for measuring this instrument is based on the fact that this operationalization gives the strongest negative correlation with concurrent and non-concurrent task variety in the first-stage regressions. See, for example Miguel et al. (2004), where instrument operationalization choice is similarly based on the strength of the first-stage regressions. Our analyses with durations shorter and longer than one year (e.g., 9 months and 15 months) reveal similar results.

[12] Our rationale is that any direct influence of vacation on surgery duration (e.g., warming up, adjustment) should disappear by the time a surgeon takes another holiday. We also consider slightly longer and shorter terms and obtain results similar in significance and effect size.

First, we examine the first-stage regression to verify the relevance of our instruments. The coefficient of *Others' Vacation* is significant and positive, whereas *Days of Vacation* is significant and negative in the first-stage regressions on *Concurrent Variety* and *Non-Concurrent Variety* with a high F-statistic and high adjusted R^2 of the corresponding models. Specifically, Table 5 shows the results of our IV approach where we use both *Other's Vacation* and *Days of Vacations* as instruments to *Non-Concurrent Variety* and *Concurrent Variety* in two stage estimation. Our instruments are significant in the first stage regression with an F-statistic well above 10 (equal to 66.03 and 78.87 respectively).

In addition, the value of the Kleibergen-Paap Wald statistic for weak instruments and the underidentification test (Kleibergen LM statistic) indicate that our instruments indeed have power and satisfy the relevance condition.[13]

After presenting the first stage, we next focus on our second-stage estimation results with the instrumental variables approach. Second-stage estimation (Model 5.3) in Table 5 provides support for both H1 and H2. Specifically, *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$), and *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$). These results are similar to our previous findings without the use of instruments.[14]

In the above IV analysis, similarly to what we discussed at the end of section 6.1, a potential concern (especially with regards to the instrument *Others' vacation*) could be that we do not specifically account for potential tenure/experience differences of surgeons and associated unobserved factors (e.g., familiarity with the hospital environment) as we do not have detailed personnel data. We therefore repeat our IV analysis only considering the subsample of operations which only include the lead surgeons who have been at the hospital from the beginning of our dataset. Our results with this analysis provide full support for H1 and H2 (Table 6).

7. Discussion and Conclusions

Professional service firms are the epitome of the increasingly knowledge-based economies of the world. While there is growing interest in the study of professional services in the management literature (Maister 1993, Hinings and Leblebici 2003, Greenwood et al. 2005, Gardner et al. 2008), the study of professional service work from an operations standpoint has been quite limited (Roth and Menor 2003, Lewis and Brown 2012). Despite the clear importance of such white-collar professions to the economies of developed countries, their operations are much less understood than the operations of blue-collar work (Hopp et al. 2009).

Learning is a critical component of white-collar work (Argote and Ingram 2000). From an operations perspective, the essential issue with regards to learning in white-collar settings is how it affects performance (Hopp et al. 2007). While past experience is generally associated with an increased

---

[13] Please note that because we have robust standard errors, we use these two tests instead of the Crag-Donald Wald F statistic and the Anderson canonical correlation LM statistic.

[14] We also consider alternative models with instruments, please see Section 6 in the appendix for details.

learning rate at the individual level (Narayanan et al. 2009, KC and Staats 2012, Staats and Gino 2012), some experiences can also have a negative effect on individual productivity (Allport et al. 1994, Lapré and Nembhard 2010, Argote and Miron-Spektor 2011, Lapré 2011). Our goal in this paper was to explore the effects of a specific kind of experience—that is, experience from performing related tasks (i.e., related variation)—on individual learning and focal productivity when the related tasks are performed concurrently vs. non-concurrently with the focal task. We find that while concurrently performing another task over time enhances the productivity of the focal task, performing another task non-concurrently reduces this productivity. We also introduce recent exposure to variety as an important moderator which amplifies the relationship between variety and productivity.

While the extant literature has recognised task variety as a potentially important driver of individuals' task performance, evidence on productivity implications of task variety have been mixed. It is only recently that researchers have started to disentangle more nuanced elements of task variety and investigate their effects on individual productivity. In an experimental study, Schilling et al. (2003) make an important distinction between related and unrelated task variation and show that the learning rate under conditions of related variation is significantly greater than under conditions of unrelated variation. In two other recent studies, Staats and Gino (2012) investigate the productivity implications of exposure to variety in the long term versus the short term, and KC and Staats (2012) introduce subtask variety (within task variety) as an important driver of individual performance. Although these studies have considerably improved our understanding of task variety's effects on individual productivity, several elements of task variety, such as the relationship between focal tasks and varied tasks and their impact on productivity, have been much less well understood. Our study sought to contribute to this line of inquiry by focusing on when the varied task has been performed with regards to the focal task. That is, we introduce the way different task are performed (concurrently vs. non-concurrently) as an important dimension to consider in understanding the effect of task variety on productivity.

Our extensive dataset, which covers a time interval of more than seven years, allowed us to investigate the influence of exposure to variety on individual learning and productivity over time. In addition, we were able to observe a variety of surgeries performed in different configurations (e.g., CABG only, other type only, CABG and other type) which are driven primarily by exogenous (i.e., medical) factors.

Surgeons have been identified as typical 21st-century knowledge workers (KC and Staats 2012) in that they experience constant learning throughout their careers. Accordingly, our results can provide useful insights for other professional service workers with regards to task allocation and timing strategies. Our findings are particularly relevant to settings that are characterized by high levels of worker discretion and control over how to conduct a variety of tasks. Our results suggest that performing other related tasks concurrently with a focal common task can improve individuals' learning and productivity over time. Our results indicate that a 20% increase in exposure to other related tasks can decrease the time an individual needs to perform the focal task by 3.69% (which, in the case of our

surgeons, translates to around 16 CABG operations per year in terms of average time savings). According to the American Heart Association (2011), a CABG costs on average $117,094. Hence, performing 16 more CABG operations may correspond to up to $1,873,504 per year in financial terms.[15] On the other hand, we observe that performing related tasks in isolation (non-concurrently) does not provide such learning and productivity-improvement opportunities. In contrast, such non-concurrent variety could be detrimental to productivity due to negative transfer effects. Our results suggest that a 20% increase in non-concurrent exposure to variety can decrease average surgery duration by 5.07%, which for our surgeons translates to around 23 CABG operations per year in terms of average time savings. Similarly, 23 more CABG operations can translate to $2,693,162 per year.

To maximize individual learning and improve productivity in common tasks, therefore, knowledge workers may consider pairing their most common task(s) with other related tasks and try performing them concurrently as much as possible. In addition, our results on the moderating role of recent exposure to variety suggest that short-term exposure to variety may matter for subsequent task productivity. That is, recent variety amplifies the respective influences of concurrent and non-concurrent variety on subsequent task productivity. Therefore, when individuals are devising strategies to improve their productivity on tasks they perform frequently, it may help to consider tasks that they have carried out both in the long term and recently.

Like all empirical studies, our study has limitations. First, although we have highly granular longitudinal data on surgery operations in terms of their type, constituents, duration, surgeon team, and other variables, we have limited information about patients' conditions before the operation. Specifically, our hospital groups patients into mild, medium, and severe categories by taking into account various clinical factors, such as their medical conditions and surgery requirements. Admittedly, this is quite a crude characterization, and there is likely variation within the three groups (e.g., some cases might be more severe than others despite being in the same severe category). Ideally, we would like to have more detailed information about patients' conditions, such as their EuroScore or Higgins score (KC et al. 2013), but this was not available in our data set. In examining the role of task variety, future work could combine detailed surgery data with more granular patient characteristics to develop more comprehensive risk and outcome measures for patients.

In addition, we should acknowledge that our data set greatly limit our ability to examine quality (e.g., mortality) implications of task variety or potential associations between surgery duration and in-hospital mortality, since our data set only includes one-month in-hospital mortality information and all mortalities come from the 'severe' patient group. Consequently, like other studies (e.g., Reagans et al. 2005, Pisano et al. 2001, Edmondson et al. 2003), we only focused on surgery duration, as we were unable to simultaneously explore other performance dimensions such as mortality. We used our limited

---

[15] One should be cautious, of course, when translating these time savings (additional operations) into financial terms, since the total cost of a CABG involves multiple elements including physician fees, hospital/room fees, the costs of drugs and medical tests, etc.

in-hospital mortality data to investigate the relationship between duration and mortality rates. Our analysis, despite its limitations, suggested that shorter durations do not come at the cost of increased mortality rates. This is also consistent with previous medical research, which in fact indicates that shorter durations could be associated with better clinical outcomes as a result of reduced risk of infections (Gaynes et al. 2001, Gibbons et al. 2011). Future research could examine how concurrent and non-concurrent exposures to variety affect other performance dimensions, such as quality. If certain task configurations provide better learning opportunities and result in higher-quality outcomes for knowledge workers, identifying and examining these would provide important contributions to operations management literature.

Also, while we believe that our particular hospital setting and our empirical approach with extensive robustness analyses considerably alleviate surgeon/patient selection concerns and we rule out many alternative explanations of our findings, in the absence of fully randomized assignment of tasks or exogenous shocks to concurrent and non-concurrent variety, our results should be interpreted with caution. Past research, for example, has highlighted potential motivational benefits from exposure to variety (Herzberg 1968, Hackman and Oldham 1976, Staats and Gino 2012), which in our case may limit the detrimental effects of non-concurrent variety. In our setting, we do not expect to observe many motivational gains: Cardiac surgery settings are high-pressure and dynamic environments (Edmondson 1999, Tucker and Edmondson 2003, Nembhard and Edmondson 2006, Wetzel et al. 2006); therefore, there is not much motivational gain to be realized as a result of exposure to variety (KC and Staats 2012). This, of course, might not be the case for settings with more repetitive and routine tasks.

Building on our insights about concurrent and non-concurrent variety and considering our work's limitations, we next discuss potential avenues for future research.

In our study, we only considered timing of task variety; however, an important discretionary responsibility of many knowledge workers involves sequencing of tasks. On this front, recent work examines drivers and implications of deviating from prescribed sequence (Ibanez et al. 2017). Additional research could explore, for example, whether the negative effects of non-concurrent variety could be mitigated in certain task-sequencing configurations.

We believe that the term 'concurrent variety', introduced in our paper, is a promising and germane concept for operations management, as it can extend existing theories of learning and knowledge transfer considerably. Many knowledge workers use discretion in deciding whether or not to perform various tasks concurrently. While the boundaries and durations of tasks are quite clear in our surgery setting, they can be less so in many other knowledge-intensive contexts. For example, a research scientist may or may not prefer concurrently working on multiple tasks or projects at the same time, each of which may also be at different stages (e.g., working on an early stage proposal draft, completing a report, revising a paper). It would be interesting to explore how learning and productivity implications of such choices may be associated with concurrent variety.

Also, in our setting, the typical duration of a focal task was approximately five hours, and the maximum number of concurrent tasks was three. Future research could consider, for example, implications of concurrent variety when a larger number of concurrent tasks are performed. As the number of concurrent tasks increases, the positive effect in the form of implicit learning from concurrent variety could potentially be reduced as a result of increased coordination issues. Future research could also explore identifying an optimal number of concurrent tasks in various knowledge-work contexts to maximize learning and productivity.

Moreover, we only focus on timing of different tasks (concurrent vs. non-concurrent), and, other than a simple robustness check with two categories involving more-related and less-related non-CABG operations, we do not consider the degree of structural similarity or difference between tasks, as we did not have a suitable way to accurately quantify this. Both the main effect of the degree of similarity between tasks on productivity and its interaction with concurrent and non-concurrent variety could be interesting to explore further. For example, the degree of structural similarity or difference between tasks may moderate the relationship between concurrent/non-concurrent variety and productivity. Also, interaction with task complexity could have an effect on concurrent learning and individual productivity.

Although surgeons are characterized as a good example of 21st century knowledge workers in the extant literature and cardiac operations provided us with a very suitable setting in which to study the role of task variety in highly cognitive and knowledge intensive yet repetitive tasks, surgery settings could be unique in terms of the nature of the tasks, their life-and-death outcomes, and surgeons' learning mechanisms. So, one has to be cautious in generalizing our findings to other professional service settings. It would be interesting to see future research on this subject and whether and to what extent our contrasting findings about concurrent and non-concurrent variety would hold in other professional service settings such as law, consulting, and research and development.

Finally, our focus in this study has been productivity implications of surgeons' learning from concurrent and non-concurrent task variety. For this reason, we are interested not in the entire surgical team but rather in individual surgeons and how they learn from task variety that takes place concurrently versus non-concurrently. This is because, while different surgery types will lead surgeons to perform different set of steps and procedures during the surgery, tasks and activities for other team members (e.g., anaesthesiologist preparing the patient, nurses providing the equipment, etc.) are more or less the same in various surgery types. So getting exposure to task variety only makes sense for surgeons. However, learning is a complex phenomenon with many individual, team-level, contextual, and organizational dynamics. Considering the recent research emphasis on the role of team-level mechanisms, such as the effect of team experience on learning and performance (Reagans et al. 2005, Huckman and Staats 2011, Avgerinos and Gokpinar 2017), future research could explore the interplay between task and team configurations and their learning implications in less hierarchical team settings with less clear task boundaries, such as management consulting or product development teams.

Acknowledgements

References

Allport D. A., G. Wylie. 1999. Task-switching: Positive and negative priming of task-set. G.W. Humphreys, J. Duncan, A.M. Treisman ed. *Attention, Space and Action: Studies in Cognitive Neuroscience*. Oxford University Press, Oxford, 273-296.

Allport, D. A., E. A. Styles, S. Hsieh. 1994. Shifting intentional set: Exploring the dynamic control of tasks. C. Umilta, M. Moscovitch, eds. *Attention and Performance XV: Conscious and Non-conscious Information Processing*. MIT Press, Cambridge, Massachusetts, 421–452.

American Heart Association. 2011. Heart disease and strike statistics 2011 Update.

Anderson, J. R. 2013. *The architecture of cognition*. Psychology Press, New York, New York.

Anderson, J. R. 1982. Acquisition of cognitive skill. *Psychological Rev*iew **89**(4) 369-406.

Argote, L. 1999. *Organizational learning: Creating, Retaining and Transferring Knowledge*. Kluwer Academic Publishers, Boston, Massachusetts.

Argote, L., E. Miron-Spektor. 2011. Organizational Learning: From Experience to Knowledge. *Organization Science* **22**(5) 1123-1137.

Argote, L., B. McEvily, R. Reagans. 2003. Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Science* **49**(4) 571-582.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organizatonal Behavior Human Decision Processes* **82**(1) 150–169.

Avgerinos E., B. Gokpinar. 2017. Team Familiarity and Productivity in Cardiac Surgery Operations: The Effect of Dispersion, Bottlenecks, and Task Complexity. *Manufacturing & Service Operations Management.***19**(1) 19-35.

Bailey, C.D. 1989. Forgetting and the learning curve: A laboratory study. *Management Science* **35**(3) 340-352.

Baltagi, B. H., and P. X. Wu. 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* **15** 814–823.

Bargh, J. A., E. L. Williams. 2006. The automaticity of social life. *Current Directions in Psychological Science* **15**(1) 1-4.

Bargh, J. A., E. L. Williams. 2007. On the automatic or nonconscious regulation of emotions. J. Gross ed. *Handbook of Emotion Regulation*. Guilford Press, New York, 429-445.

Bargh, J. A., K. L. Schwader, S. E. Hailey, R. L. Dyer, E. J. Boothby. 2012. Automaticity in social-cognitive processes. *Trends in Cognitive Science.* **16**(12) 593-605.

Ben-Zeev, T., J.R. Star. 2001. Spurious correlations in mathematical thinking. *Cognition and Instruction*. **19**(3) 253-275.

Boh, W. F., S. A. Slaughter, J. A. Espinosa. 2007. Learning from experience in software development: A multilevel analysis. *Management Science* **53**(8) 1315–1331.

Bransford J.D., J.J. Franks. 1976. Toward a framework of understanding learning. *Psychology of Learning and Motivation.* **10** 93-127

Breusch, T. S., A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica, Econometric. Society* **47**(5) 1287-1294.

Cellier, J. M., H. Eyrolle. 1992. Interference between switched tasks. *Ergonomics* **35**(1) 25–36.

Ch'ng, S. L., A. D. Cochrane, R. Wolfe, C. Reid, C. I. Smith, J. A. Smith. 2015. Procedure-specific cardiac surgen volume associated with patient outcome following valve surgery, but not isolated CABG surgery. *Heart, Lung and Circulation*. **24**(6) 583-589.

Clark, J. R., R. Huckman. 2012. Broadening focus: Spillovers, complementarities and specialization in the hospital industry. *Management Science* **58**(4) 708-722.

Craik, F.I. M., E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology* **104**(3) 268-294.

Durbin J. 1954. Errors in Variables. *Review of the International Statistics Institute.* **22**(1/3) 23-32.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* **44**(2) 350-383.

Edmondson, A. C., A. B. Winslow, R. M. J. Bohmer, G. P. Pisano. 2003. Learning how and learning what: Effects of tacit and codified knowledge on performance improvement following technology adoption. *Decision Sci*ences **34**(2) 197-224.

Egelman C.D, D. Epple, L. Argote, E.R.H. Fuchs. 2016. Learning by Doing in Multiproduct Manufacturing: Variety, Customizations, and Overlapping Product Generations. *Management Science* **63**(2) 405-423.

Gardner, H. K., Anand, N., & Morris, T. 2008. Chartering new territory: Diversification, legitimacy, and practice area creation in professional service firms. *Journal of Organizational Behavior* **29**(8) 1101–1121.

Gaynes, R. P., D. H. Culver, T. C. Horan, J. R. Edwards, C. Richards, J. S. Tolson, and the National Nosocomial Infections Surveillance System. 2001. Surgical site infections (SSI) rates in the United States, 1992-1998: The national nosocomial infections surveillance system basic SSI risk index. *Clinical Infectious Diseases* **33**(Suppl 2) 69-77.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* **7**(2) 155–170.

Gentner, D. 1989. The mechanisms of analogical learning. S. Vosniadou, A. Ortony ed. *Similarity and analogical reasoning.* Cambridge University Press, New York 199-241.

Gentner, D., A.B. Markman. 1997. Structure-mapping in analogy and similarity. *American Psychologist.* **52**(1) 45-56.

Gentner, D., J. Colhoun. 2010. Analogical processes in human thinking and learning. B. Glatzeder, V. Goel, A. Muller. ed. *Towards a Theory of Thinking.* Springer-Verlag, Berlin 35-48.

Gibbons, C., J. Bruce, J. Carpenter, A. P. Wilson, J. Wilson, A. Pearson, D. L. Lamping, Z. H. Krukowski, B. C. Reeves. 2011. Identifications of risk factors by systematic review and development of risk-adjusted models for surgical site infection. *Health Technological Assessment* **15**(30) 1-156.

Gick, M. L., K. L. Holyoak. 1987. The cognitive basis of knowledge transfer. S. M. Cormier, J. D. Hagman eds. *Transfer of Learning: Contemporary Research and Applications.* Academic Press, New York.

Gladstein, D. L. 1984. Groups in context: A model of task group effectiveness. *Administrative Science. Quarterly* **29**(4) 499-517.

Globerson, S., N. Levin, A. Shtub. 1989. The impact of breaks on forgetting when performing a repetitive task. *IIE Transactions* **21**(4) 376-381.

Graham, C., R. Gagne. 1940. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *Journal of Experimental Psychology* **26**(3) 251-280.

Greenwood, R., Li, S. X., Prakash, R., & Deephouse, D. L. 2005. Reputation, diversification, and organizational explanations of performance in professional service firms. *Organization Science* **16**(6) 661–673.

Grierson, L., M. Melnyk, N. Jowlett, D. Backstein, A. Dubrowski. 2011. Bench model surgical skill training improves novice ability to multitask: A randomized controlled study. J. D. Wetswood, S.W. Westwood, L. Fellander-Tsai, R.S. Haluck, H.M. Hoffman, R.A. Robb, S. Senger, K.G. Vosburgh ed. *Medicine meets virtual reality 18.* IOS Press, Netherlands 192-198.

Hackman, J. R. 2002. *Leading Teams: Setting the Stage for Great Performances.* Harvard Business Press, Boston, Massachusetts.

Hackman, J. R., G. R. Oldham. 1976. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance* **16**(2) 250-279.

Hausman J. A. 1978. Specification tests in econometrics. *Econometrica* **46**(6) 1251-1271.

Herzberg, F. 1968. One more time: How do you motivate employees? *Harvard Business Review* **46**(1) 53–62.

Hinings, C. R., H. Leblebici. 2003. Introduction: Knowledge and professional organizations. *Organization Studies* **24**(6) 827–830.

Holyoak K.J, K. Koh. 1987. Surface and structural similarity in analogical transfer. *Memory and Cognition* **15**(4) 332-340.

Holyoak, K.J., P.R. Thagard. 1995. Mental leaps: Analogy in creative thought. MIT Press, Cambridge, Massachusetts.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Production and Operations Management* **18**(1) 1-32.

Hopp, W. J., S. M. R. Iravani, G. Yuen. 2007. Discretionary task completion: A key difference between white-collar and blue collar work systems. *Management Science* **53**(1): 61–77.

Huckman, R. S. 2003. The utilization of competing technologies within the firm: Evidence from cardiac procedures. *Management Science* **49**(5) 599–617.

Huckman, R. S., G. P. Pisano. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science* **52**(4) 473–488.

Huckman, R. S., BR. Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management.* **13**(3) 310-328.

Iacus S. M., King G., Porro G. 2011. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association* **493**(106) 345-361.

Ibanez M., J. R. Clark, R. S. Huckman, B. R. Staats. 2017. Discretionary Task Ordering: Queue Management in Radiological Services. *Management Sci*ence Forthcoming.

Kahol, K., A. Ashby, M. Smith, J. Ferrara. 2010. Quantitative evaluation of retention of surgical skills in simulation. *Journal of Surgical Education* **67**(6) 421-426.

KC, D. S., BR. 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.S., B.R. Staats, F. Gino. 2013. Learning from My Success and from Others' Failure: Evidence from Minimally Invasive Cardiac Surgery. *Management Science* **59**(11) 2435-2449.

Kimiyoshi J. B.S. Kobayashi, J.A. Williams, L.U. Nwakanma, E.S. Weiss, V.L. Gott, W.A. Baumgartner, J.V. Conte. 2009. EuroSCORE Predicts Short- and Mid-Term Mortality in Combined Aortic Valve Replacement and Coronary Artery Bypass Patients. *Journal of Cardiac Surgery* **24**(6) 637-643.

Koechlin, E., G. Basso, P. Pietrini, S. Panzer, J. Grafman. 1999. The role of the anterior prefrontal cortex in human cognition. *Nature* **399**(6732) 148-151.

Kouchoukos, N.T., E.H. Blackstone, F.L. Hanley, J.K. Kirklin. 2012. *Cardiac Surgery* Vol. 1. Kirklin/Barratt Boyes. Elsevier Saunders, Philadelphia, Pennsylvania.

Koechlin, E., A. Hyafil. 2007. Anterior prefrontal function and the limits of human decisionmaking. *Science* **318**(5850), 594-598.

Lance, C. E., A. G. Parisi, W. R. Bennett, M. S. Teachout, D. L. Harville, M. L. Welles. 1998. Moderators of skill retention interval/performance decrement relationships in eight U.S. Air Force enlisted specialties. *Human Performance* **11**(1) 103-123.

Langer, E. J. 1989. *Mindfulness*. Addison-Wesley Publications, Reading, Massachusetts.

Lapré, M. A. 2011. Inside the learning curve: Opening the black box of the learning curve. Jaber M. Y., ed. *Learning Curves: Theory, Models, and Applications*. CRC Press, Boca Raton, Florida 23–35.

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

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

Lassaline, M.E., G.L. Murphy. 1998. Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* **24**(1) 144-160.

Lee, T. D., D. Simon. 2004. Contextual Interference. Williams, A. M., N. J. Hodgeseds ed. *Skill Acquisition in Sport: Research, Theory and Practice*. Routledge, London, 29-44.

Lerner, J. S., D. A. Small, G. Loewenstein. 2004. Heart strings and purse strings: Carry-over effects of emotions on economic transactions. *Psychological Science* **15**(5) 337-341.

Lewis, M. A., A. D. Brown. 2012. How different is professional service operations management? *Journal of Operations Management* **30**(2) 1-11.

Loewenstein, J., L. Thompson, D. Gentner. 1999. Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review* **6**(4) 586-597.

Loewenstein, G., J. S. Lerner. 2002. The role of affect in decision making. Davidson, R., K. Scherer, H. Goldsmith eds. *Handbook of Affective Science*. Oxford University Press, New York, 619-642.

Loewenstein, J., L. Thopson, D. Gentner. 2003. Analogical learning in negotiation teams: Comparing cases promotes learning and transfer. *Academy of Management Learning and Education* **2**(2) 119-127.

Maister, D. H. 1993. *Managing the Professional Service Firm*, The Free Press, New York.

Maskarinec, A. S., C. P. Thompson. 1976. The within-list distributed practice effect: Tests of the varied context and varied encoding hypothesis. *Memory and Cognition* **4**(6) 741–746.

Markman, A.B., D. Gentner. 1993. Structural alignment during similarity comparisons. *Cognitive Psychology* **25**(4) 431-467.

McKinsey Global institute. 2012. The social economy: Unlocking value and productivity through social technologies. November 2012.

Medin, D.L., R.L. Goldstone, D. Gentner. 1993. Respects for similarity. *Psychological Review* **100**(2) 254-278.

Miguel, E., S. Satyanath, E. Sergenti. 2004. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy* **112**(4), 725-753.

Monsell, S. 2003. Task switching. *Trends in Cognitive Sci*ence **7**(3) 134-140.

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

Nembhard, D. A., M. V. Uzumeri. 2000. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* **25**(4) 315-326.

Nembhard, I. M., A. C. Edmondson. 2006. Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behav*ior **27**(7) 941-966.

Novick, L.R. 1988. Analogical transfer, problem similarity, and experitise. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**(3) 510-520.

Pisano, G. P., R. M. J. Bohmer, A. C. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci*ence **47**(6) 752-768.

PwC. 2012. What are professional services? Retrieved August 15, 2014, http://www.pwc.co.uk.

Ramdas, K., K. Saleh, S. Stern, Liux H. 2017. Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science* Forthcoming.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Sci*ence **51**(6) 869-881.

Reber, A. 1989. Implicit learning and tacit knowledge. *Journal Experimental Psychology* **118**(3) 219–235.

Reder, L. M. 1982. Plausibility judgements versus fact retrieval: Alternative strategies for sentence verification. *Psychology Review* **89**(3) 90-137.

Reznick, R. K., H. MacRae. 2006. Teaching surgical skills-changes in the wind. *The New England Journal of Medicine* **355** 2664-2669.

Reed, S.K. 1987. A structure-mapping problem for words problems. *Journal of Experimental Psychology: Learning, Memory and Cognition* **13**(1) 124-139.

Ross, B.H. 1987. This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory and Cognition* **13**(4) 629-639.

Roth, A. V., & Menor, L. J. 2003. Insights into service operations management: a research agenda. *Production and Operations Management.* **12**(2) 145-164.

Saxena A., D. Dinh, J. A. Smith, C. M. Reid, G. C. Shardey, A. E. Newcomb. 2013. Excellent short- and long-term outcomes after concomitant aortic valve replacement and coronary artery bypass grafting performed by surgeons in training. *The Journal of Thoracic and Cardiovascular Surgery* **145**(2) 334-340.

Schilling, M. A., P. Vidal, R. E. Ployhart, A. Marangoni. 2003. Learning by doing something else: Variation, relatedness, and the learning curve. *Management Sci*ence **49**(1) 39-56.

Shtub A., N. Levin, S. Globerson. 1993. Learning and forgetting industrial skills: An experimental model. *International Journal of Human Factors in Manufacturing* **3**(3) 293-305.

Spencer, F.C. 1978. Teaching and measuring surgical techniques-the technical evaluation of competence. *Bulletin of the American College of Surgeons* **63**(3) 9-12.

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

Thurstone, L. L. 1919. The learning curve equation. *Psychological Monographs: General and Applications* **26**(3) 1-51.

Tucker, A. L., A. C. Edmondson. 2003. Why hospitals don't learn from failures: organizational and psychological dynamics that inhibit system change. *California Management Rev*iew **45**(2) 55-72.

Tucker, A. L., I. M. Nembhard, A. C. Edmondson. 2007. Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Sci*ence **53**(6) 894-907.

Tulving, E., D.M. Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* **80**(5) 352-373.

Tversky, A. 1977. Features of similarity. *Psychological Review* **84**(4) 327.

Von Nordenflycht, A. 2010. What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. *Academy of Management Review* **35**(1) 155-174.

Wetzel, C. M., R. L. Kneebone, M. Woloshynowych, D. Nestel, K. Moorthy, J. Kidd, A. Darzi. 2006. The effects of stress on surgical performance. *The American Journal of Surgery* **191**(1) 5-10.

Wu D. M. 1973. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* **41**(4) 733-750.

Wulf, G., R. A. Schmidt. 1997. Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**(4) 987–1006.

Zollo, M., J. J. Reuer. 2010. Experience spillovers across corporate development activities. *Organization Science* **21**(6) 1195-1212.

Table 2. Descriptive Statistics for Lead Surgeons

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Duration	5.681	0.176	4.094	6.579	1													
2. Recent Non-Concurrent Variety	0.741	0.633	0	1.386	0.011**	1												
3. Non-Concurrent Variety	4.999	0.983	0	5.820	0.121*	0.598**	1											
4. Recent Concurrent Variety	0.505	0.601	0	1.792	-0.019*	0.494**	0.274**	1										
5. Concurrent Variety	4.904	0.984	0	5.489	0.092**	0.487**	0.584**	0.208**	1									
6. Recent Non-Concurrent Variety x Non-Concurrent Variety	4.827	1.954	0	7.077	0.023*	0.828**	0.611**	0.342**	0.611**	1								
7. Recent Concurrent Variety x Concurrent Variety	4.184	1.997	0	6.956	-0.037*	0.389**	0.447**	0.884**	0.471**	0.558**	1							
8. Focal Experience	5.978	1.106	0	6.477	0.052**	0.354**	0.578**	0.254**	0.689**	0.421**	0.289**	1						
9. Team Size	1.617	0.130	0.693	2.079	0.117**	-0.097*	-0.043**	-0.016	0.052*	-0.311**	-0.273	0.049**	1					
10. Average Individual Experience	6.085	1.100	0	7.701	0.103**	0.101*	0.533**	0.041**	0.623**	0.477**	0.293**	0.686**	0.069**	1				
11. Severe	0.043	0.203	0	1	0.036*	0.035*	0.011*	-0.013	-0.061**	-0.013	-0.008	-0.048**	0.024	-0.064**	1			
12. Mild	0.291	0.454	0	1	-0.061**	-0.011	-0.014+	-0.018	-0.029	-0.011	0.005	-0.035*	-0.024	-0.047**	-0.136**	1		
13. Age	64.931	11.988	8	88	0.016	-0.012	-0.022	-0.007	-0.020	-0.020	-0.016	-0.016	0.034+	0.025	0.011	-0.001	1	
14. Male	0.738	0.440	0	1	0.004	-0.002	-0.020	-0.012	0.023	-0.017	-0.026	-0.027	-0.006	-0.029	0.005	-0.013	-0.084**	1

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Logged values of all variables except Severe, Mild, Age and Male

--Concurrent Variety x Recent Concurrent Variety | | | | | -0.025* |
					(0.011)
Focal Experience	-0.122**	-0.135**	-0.126**	-0.129**	-0.126**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Team Size	0.095**	0.128**	0.119**	0.119**	0.119**
	(0.025)	(0.033)	(0.033)	(0.033)	(0.033)
Average Individual Experience	-0.029**	-0.027**	-0.026**	-0.025**	-0.025**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Severe	0.031*	0.030*	0.027*	0.028*	0.028*
	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
Mild	-0.013*	-0.013*	-0.012*	-0.011+	-0.011+
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.004	0.005	0.003	0.003	0.003
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	3.970**	3.917**	4.036**	4.122**	4.132**
	(0.110)	(0.115)	(0.115)	(0.117)	(0.117)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.145	0.150	0.162	0.165	0.168
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 4. Regression of Task Variety on Surgery Duration for Lead Surgeons with Subsample

Variable	Duration			
	Model: (4.1)	(4.2)	(4.3)	(4.4)
Non-Concurrent Variety	0.133**	0.195**	0.207**	0.202**
	(0.048)	(0.055)	(0.056)	(0.057)
Concurrent Variety		-0.193*	-0.190*	-0.183*
		(0.056)	(0.057)	(0.058)
Recent Non-Concurrent Variety			0.030	0.025
			(0.025)	(0.028)
Recent Concurrent Variety				-0.012
				(0.023)
Non-Concurrent Variety x Recent Non-Concurrent Variety			0.016**	0.016**
			(0.004)	(0.005)
Concurrent Variety x Recent Concurrent Variety				-0.013*
				(0.006)
Focal Experience	-0.104**	-0.081**	-0.083**	-0.084**
	(0.019)	(0.019)	(0.019)	(0.019)
Team Size	0.160**	0.158**	0.155*	0.155**
	(0.036)	(0.036)	(0.036)	(0.036)
Average Individual Experience	-0.011*	-0.016**	-0.017**	-0.016**
	(0.005)	(0.005)	(0.005)	(0.005)
Severe	0.025+	0.026+	0.026+	0.025+
	(0.014)	(0.015)	(0.015)	(0.014)
Mild	-0.010	-0.008	-0.008	-0.007
	(0.007)	(0.007)	(0.007)	(0.007)
Age	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.000	-0.001	0.002	-0.001
	(0.007)	(0.007)	(0.009)	(0.007)
Constant	4.110**	4.146**	4.188**	4.200**
	(0.139)	(0.138)	(0.142)	(0.142)
Observations (N)	2308	2308	2308	2308
Adjusted R^2	0.140	0.163	0.165	0.166
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 5. Regression of Task Variety on Surgery Duration with Instrumental Variables Approach

Variable	Non-Concurrent Variety	Concurrent Variety	Duration
	Model: (5.1)	(5.2)	(5.3)
Non-Concurrent Variety			0.297**
			(0.029)
Concurrent Variety			-0.195**
			(0.036)
Other's Vacation	0.00002**	0.00002**	
	(0.00000)	(0.00000)	
Days of Vacation	-0.005**	-0.006**	
	(0.001)	(0.001)	
Focal Experience	0.606**	0.732**	-0.263**
	(0.013)	(0.013)	(0.044)
Team Size	-0.648**	-0.669**	0.187**
	(0.062)	(0.063)	(0.059)
Average Individual Experience	0.079**	0.095**	-0.065**
	(0.012)	(0.012)	(0.018)
Severe	-0.022	-0.028	0.025*
	(0.035)	(0.036)	(0.012)
Mild	0.007	0.019	-0.009
	(0.016)	(0.016)	(0.010)
Age	0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)
Male	0.004	-0.007	-0.006
	(0.016)	(0.016)	(0.010)
Constant	0.409**	-0.142	4.647**
	(0.110)	(0.112)	(0.202)
Observations (N)	3273	3273	3261
Adjusted R^2	0.419	0.436	0.147
Day of the Week Fixed Effect	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively

Table 6. Regression of Task Variety on Surgery Duration with Instrumental Variables Approach with Subsample

Variable	Non-Concurrent Variety	Concurrent Variety	Duration
	Model: (6.1)	(6.2)	(6.3)
Non-Concurrent Variety			0.183**
			(0.050)
Concurrent Variety			-0.109*
			(0.049)
Other's Vacation	0.00001**	0.00001**	
	(0.00000)	(0.00000)	
Days of Vacation	-0.004**	-0.005**	
	(0.001)	(0.001)	
Focal Experience	0.759**	0.890**	-0.143**
	(0.012)	(0.013)	(0.040)
Team Size	-0.856**	-0.879**	0.122*
	(0.055)	(0.057)	(0.058)
Average Individual Experience	0.037**	0.054**	-0.035**
	(0.010)	(0.011)	(0.009)
Severe	-0.036	-0.043	0.022+
	(0.032)	(0.033)	(0.012)
Mild	0.005	0.019	-0.004
	(0.014)	(0.014)	(0.008)
Age	0.0002	0.00006	0.00003
	(0.001)	(0.001)	(0.00006)
Male	0.008	-0.001	-0.006
	(0.014)	(0.015)	(0.010)
Constant	-0.217**	0.033	4.637**
	(0.100)	(0.104)	(0.158)
Observations (N)	2315	2315	2308
Adjusted R^2	0.387	0.399	0.134
Day of the Week Fixed Effect	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes

+, * and ** denote significance at 10%, 5% and 1% levels respectively