Team and Task Dynamics in Healthcare and Professional Service Operations

by

Emmanouil Avgerinos

A thesis submitted for the degree of
Doctor of Philosophy in
Management Science and Innovation

January 2016

UCL School of Management
University College London

Declaration

I, Emmanouil Avgerinos confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, January 2016

Abstract

This thesis examines how team, task and workforce dynamics affect performance on healthcare and professional service operations. I use operations management and organizational domains to develop theories and employ econometric models to better understand knowledge intensive environments. In the first chapter I use data from Coronary Artery Bypass Graft operations in order to examine the way exposure to related variety can affect individual learning. Specifically, I introduce timing as a new dimension on the effect of related variety on individual productivity on a focal task and show that exposure to variety can have differentiated effects on individual productivity based on different mechanisms. My findings suggest that concurrent exposure with the focal task has a positive effect whereas non-concurrent one has a negative effect on individual focal productivity. I also introduce recent concurrent and non-concurrent exposures as moderating factors on the effect of long-term concurrent and non-concurrent exposures respectively on individual learning.

In the second chapter I focus on cardiac surgery teams and examine the effect of team allocation on their productivity. Specifically, I introduce new familiarity related concepts and ways on how past common experiences among team members can affect team productivity. Next, I divide average team familiarity into two components: One gained from complex and one gained from simpler tasks and show their differentiated effects on team productivity. I also investigate the way average team familiarity interacts with task complexity.

In the final chapter, I use a dataset from England's National Health Service (NHS)'s 111 non-emergency helpline in order to investigate the effect of non-clinical labor mix on efficiency and quality of patient service. My results indicate that while non-clinical workforce increases the efficiency of patient service, it may lead to new inefficiencies through misuse of critical resources and may reduce the quality outcome of the patient service.

Acknowledgements

I am deeply grateful to my supervisor and academic mentor Bilal Gokpinar for his invaluable guidance and continuous support in every single step of this thesis. I am also thankful to Bert De Reyck for his contribution to my academic progress and development.

Special thanks to Ioannis Fragkos, Enrico Forti, Kremena Slavova, Sebastiano Massaro and Chia-Yu Kou-Barrett for their advice over the first years of my PhD studies. I am also grateful to Rouba Ibrahim, Bart Vanneste, James Berry, Mihaela Stan, Sarah Harvey, Nicos Savva, Martin Kilduff and Simcha Jong for their useful comments on the first two chapters of my thesis.

I also owe my deep gratitude to Iason Bakogiannis for his help during my PhD studies and to Ivan Campos for his support.

I would also like to thank the Foundation of Education and European Culture (IPEP) and its President Nikos Trichas for their financial support during the first years of my PhD studies.

Last but not least, I would like to thank most of all my family for their invaluable patience and support over all these years.

Table of Contents

Chapter four - The Role of Non-clinical Workforce on Efficiency and Quality of Patient Service: Evidence from NHS Medical Helpline

Chapter five -Conclusions

List of Figures

List of Tables

To my family

Chapter one – Introduction

The term “white-collar workers” refers to individuals who work in knowledge intensive environments and perform non-routine and intellectual tasks. Despite the increasing significance of their role to the economy, they have received little attention in Operations Management research, with compare to the traditional blue-collar workers that perform physical tasks. Specifically, little is understood about the principles of how an operation consisting of white-collar individuals should be structured and performed in order to achieve high performance. My work focuses on developing theories and empirically testing them in knowledge intensive environments such as healthcare settings since healthcare employees represent typical knowledge workers. This thesis contributes to the nascent literature of empirical research in professional service individuals, teams and organizations and aims to introduce novel ways of increasing their performance.

First, I focus on the effect of exposure to related variety on individual productivity. Despite the intellectual nature of white-collar tasks, knowledge workers are often assigned to perform different but related tasks. Another important feature of white-collar workers is that they tend to have discretion in deciding when to perform a specific task or how to perform it (independently or concurrently with other related ones). In this chapter I use data from Coronary Artery Bypass Graft operations in order to develop and test a model for knowledge transfer based on the way exposure to related variety can affect individual learning over time. Specifically, I first introduce timing as a new dimension on the effect of related variety on individual productivity on a focal task and show that exposure to variety can have differentiated effects on individual productivity based on different mechanisms (concurrently vs. non-concurrently with the focal task). My findings suggest that concurrent exposure has a positive effect whereas non-concurrent one has a negative effect on individual focal productivity. In addition, I introduce recent concurrent and non-concurrent exposures as moderating factors on the effect of long-term concurrent and non-concurrent exposures respectively on individual learning.

Second, I investigate fluid teams and how team familiarity can affect their productivity. Fluid teams refer to teams that have no permanent membership. They operate as a team for a specific task, have clear hierarchies and responsibilities and dissolve after performing the assigned task. One key management tool for increasing team productivity is team allocation since there is wide consensus that previous shared experience (i.e., team familiarity) among team members is beneficial for team productivity. In this work, I use cardiac surgery data in order to extend the effect of average team familiarity by introducing new familiarity related metrics and ways on how past common experiences among team members can affect team productivity. Specifically, I introduce “team familiarity dispersion”, which captures the different levels of familiarity among pairs within the same team, and show that it can decrease team productivity. Next, inspired by the concept of bottleneck in blue-collar settings, I create an analogous construct “bottleneck pair” for fluid teams, which I define as a pair of the team with very low familiarity compared to the average level of team familiarity and show how it can lead to lower team productivity. Next, I divide average team familiarity into two components: One gained from complex and one gained from simpler tasks and show that the former has a higher beneficial impact on team productivity that the latter one. I also investigate the way average team familiarity interacts with task complexity and find that the beneficial effect of the former is more prominent when performing more complex tasks.

Finally, I examine the effect of non-clinical workers on the performance of healthcare organizations. Such organizations increasingly rely on a mix of clinical and non-clinical health personnel in providing innovative healthcare services such as medical helplines. While these can offer significant cost and patient access advantages, determining the right mix of health personnel is a major challenge in such settings. In this study, making use of a dataset (i.e., England’s National Health Service (NHS)’s new 111 non-emergency helpline), I investigate the effect of non-clinical labor mix on efficiency and quality of patient service. My results indicate that while non-clinical workforce increases the efficiency of patient service by reducing abandoned calls, it may lead to new inefficiencies through misuse of critical resources such as unnecessary ambulance dispatches and also reduce the quality outcome of the patient service.

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

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.

2.1. 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 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), they also demonstrate several key features from an operations standpoint in terms of the way their employees perform their work.

First, the majority of the work performed by professional service workers 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 a similar set of activities; and surgeons perform in the operating room by following a certain set of procedures. While most activities and tasks may be quite similar from one job to another, one still observes high cognitive activity among workers, presumably due to the variation in work content across tasks (e.g., differences between consulting projects, between surgeries, or legal cases). Hopp et al. (2009) consider this key observation in their classification of white-collar work: they call these (e.g., consulting, 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). 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—an important operational question is the following: 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 seeks to address these questions by focusing on the way various tasks are performed. We distinguish between concurrent and non-concurrent exposure to variety and examine the 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 the way (i.e., 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.

Specifically, 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. Concurrent variety is beneficial for productivity because it 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. Non-concurrent variety, on the other hand, results in reduced productivity through cognitive interference (Sarason and Pierce 1996), invalid generalizations (Gick and Holyoak 1987), and priming (Allport et al. 1994). In addition, our results suggest that short-term exposure to variety amplifies the influence of the long-term ones 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 through better allocation of a variety of tasks they perform, we should emphasize an important empirical consideration. Because many 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. In order to test the hypotheses developed in the present paper, ideally one needs 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 instead are 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 our setting is ideal in which to investigate the role of task variety on productivity.

Our study 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. (2003)'s 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 study contributes to the growing literature on task variety and productivity by introducing a new dimension of variety which has not been considered before—that is, concurrent and non-concurrent exposure. In a similar study, Staats and Gino (2012) focused on a single current task, and studied when different tasks take place (on the same day or in the past) with respect to that current task, and examine how these affect current task performance. We, however, concentrate on a common focal task being performed over time (i.e., CABG), and 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 their 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.

2.2. Setting

Our setting is the cardiac unit of a private hospital in Europe, and our dataset 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 2.4 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 dataset 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 previously, 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.

In this setting, we identify CABG as the focal surgery 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.

2.3 Literature Review and Theory Development

2.3.1 Exposure to Variety, Learning, and Productivity

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 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 is 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. Task variation can be established by performing either related tasks (related variation) or unrelated tasks (unrelated variation) (see Schilling et al. 2003 for a discussion of this distinction). Our study only considers related variation, since most professional service workers perform highly related tasks as part of their jobs (e.g., preparing various legal cases or performing related surgeries). Consequently, throughout this paper, when we say "other task", we refer to a task that is different from but related to the focal one.

Organizational research that examines the performance implications of exposure to task variety at the individual, team, or organizational level has found mixed results.

Clark and Huckman (2012), for example, found no evidence that related activities generated positive spillovers on the focal activity within operating units in different hospitals. Narayanan et al. (2009) found that variety had an inverted U-shaped relationship with individual productivity in an offshore software support services company. On the firm level, Haleblian and Finkelstein (1999) suggested that prior related experience has a U-shaped relationship with the focal experience in the context of acquisitions. Finally, 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.

Research in cognitive psychology shows that the effects of experience in related tasks depend-among other factors-on the degree of similarity between related and focal tasks and the type of knowledge that is transferred between the two tasks. Repetitive tasks which represent a significant level of similarity and require simple skills are likely to present positive effects and promote individual learning (Monsell 2003, Zollo and Reuer 2010).

There is also a significant body of literature that argues that task variety enhances learning through successful transfer of knowledge between the focal and related activity. The more similar these activities are, the higher the probability is of a successful knowledge transfer and application (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. Boh et al. (2007) use data from the software industry to show that, on a team and organizational level, exposure to related systems is more beneficial for performance than specialization. Schilling et al. (2003) show that related variety can improve the learning rate of students playing different versions of a game. Staats and Gino (2012) use data from a Japanese bank to show that variety promotes workers' productivity in the long run.

On the other hand, researchers have also suggested that prior learning in related tasks can have a negative effect (Gick and Holyoak 1987, Zollo and Reuer 2010). Gick and Holyoak (1987) call this "negative transfer effect" and demonstrate its existence at the individual level. Holyoak (1985) shows that when there is a difference between an individual's perception of similarity and the actual similarity of the tasks, a negative transfer effect of knowledge can appear through invalid generalizations. Holland (1986) and Holyoak (1985) also introduced the term "brittleness", which is defined as an individual's inability to transfer knowledge from previous related experience to new tasks. Cohen and Bacdayan (1994) found that individuals who played a card game had performance difficulties when the rules were slightly changed.

Finally, researchers have argued that task variety could be distracting and eventually detrimental to individual learning (Allport et al. 1994, Monsell 2003). Exposure to variety can have a negative effect on the focal task, since it can entail switching costs, warm-up periods, residual costs, and mixing costs (Monsell 2003). Specifically, shifting between different tasks can be distracting (Schultz et al. 2003, Staats and Gino 2012) and the individual may therefore need a certain amount of time to reconfigure the new task (Rogers and Monsell 1995, De Jong 2000, Rubinstein et al. 2001) or adopt the task-specific behaviour (Allport and Wylie 1999, Monsell 2003, Waszak et al. 2003). Task variety could therefore lead to lower productivity.

2.3.2 Concurrent and Non-Concurrent Exposure to Variety

Our study 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 by arguing that the impact of task variety on learning and productivity critically depends on the way related tasks are performed. Specifically, we suggest that the impact of exposure to a related task independently (non-concurrently) versus concurrently with the focal task determines the level of successful knowledge transfer and associated learning from the related task to the focal task. 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 happen concurrently with the focal task.

We argue that a key mechanism for individuals in transferring task related knowledge to subsequent tasks is whether related tasks are performed at the same time or at a different time to the focal task. That is, non-concurrent task variety could inhibit task-related 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. 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.

2.3.3 Non-Concurrent Exposure to Variety

Previous studies on the positive effects of exposure to task variety on performance highlight two basic mechanisms. These include positive knowledge transfer as a result of exposure to variety (Monsell 2003, Zollo and Reuer 2010) and motivational benefits due to task variety (Herzberg 1968, Hackman and Oldham 1976). Because cardiac surgeries are highly complex tasks that demand a combination of cognitive and motor skills (Schaverian 2010), we do not expect to see straightforward knowledge transfer from task variety, particularly when other tasks are performed non-concurrently. Since even minor differences between operations can have a significant impact on their outcome, we do not expect to observe a positive knowledge transfer from such exposure. Similarly, because cardiac surgery settings are high-pressure and dynamic environments (Edmondson 1999, Tucker and Edmondson 2003, Nembhard and Edmondson 2006, Wetzel et al. 2006), there is not much motivational gain to be realized as a result of exposure to variety (KC and Staats 2012). Indeed, Ch'ng et al. (2015) find no benefit of task variety for cardiac surgeons. Specifically, their results indicate that performing more valve operations does not improve surgeons' performance on CABG operations and vice versa.

Scholars in medical research consider surgery operations to require twenty-five percent technical and seventy-five percent decision-making skills (Spencer 1978, Grierson et al. 2011). While non-concurrent variety may help surgeons learn technical skills, we argue that non-concurrent variety provides more costs than benefits for the decision-making skills that are more important in the dynamic and high-pressure surgery environment. These costs come about for several reasons.

First, non-concurrent exposure to variety (e.g., performing different surgery types that do not include a CABG) is likely to lead to cognitive interference (Sarason and Pierce 1996). The individual has to devote cognitive resources to completing the different task, which decreases the resources available the next time the individual performs the focal task (Wylie and Allport 2000, Waszak et al. 2003, Staats and Gino 2012, KC and Staats 2012). This effect will be more significant over time as the number of different surgery types performed increases, resulting in a decrease in available cognitive resources. Surgeons tend to experience high cognitive load associated with developing the mental structures that organize the procedural steps for surgeries (Skaugset et al. 2015). Furthermore, increased cognitive interference may lead to increased stress (Wetzel et al. 2006). Surgeons tend to suffer from stress, especially in high-pressure environments such as cardiac surgeries, and stress decreases their productivity (Balch and Shanafelt 2011, Orri et al. 2015). Overall, non-concurrent task variety will lead to reduced productivity by consuming additional cognitive resources which are highly critical but limited (Cuschieri et al. 2001, Youngson 2000) and by increasing stress.

Second, development of implicit memory (i.e., priming) due to non-concurrent variety can impair decision making in a subsequent focal task. Individuals evoke different sets of actions in response to a stimulus when performing another task (Allport et al. 1994, Allport and Wyllie 1999, Wyllie and Allport 2000, Waszak et al. 2003, Staats and KC 2012), and researchers have shown that unconscious response from past stimuli does occur among doctors and medical staff (Loewenstein and Lerner 2002, Bargh and Williams 2007). This unconscious response not only interferes with and lengthens the focal task (Allport et al. 1994), but also can lead to suboptimal decisions and action sets in the focal surgery (i.e., CABG). That is, although each surgery type is unique and requires its own distinct set of steps (Reznick and MacRae 2006), priming due to non-concurrent variety may result in surgeons responding to different but related surgery types with a number of action sets, some of which may be inappropriate and even detrimental for the focal task. Also, the number of action sets that surgeons automatically adopt when facing a related experience will increase over time, and this may in turn decrease surgeons' productivity when they perform the focal task.

Third, while most cardiac surgery types follow similar procedures and principles, even minor inherent differences in these highly critical tasks may lead to significant practical differences in the operating room. There is, therefore, an increased probability that surgeons may misjudge the level of similarity between the focal and other surgery types, which may result in misguided generalizations and hence lower productivity. Any difference between cognitive perceptions and the actual level of knowledge can lead to decreased productivity (Hollyoak 1985, Zollo and Reuer 2010). So, as exposure to variety that does not include the focal surgery type increases (i.e., non-concurrent variety), invalid generalizations can create a negative effect of prior learning from related tasks. When the individual subsequently performs the focal task, a negative knowledge transfer will take place (Gick and Holyoak 1987), which will decrease 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.*

2.3.4 Concurrent Exposure to Variety

We next examine the productivity implications of performing other tasks in conjunction with the focal task. When another surgery is performed concurrently with the focal one, the dangers of invalid generalizations over time and negative learning transfer are significantly reduced. This is because the individual is also performing the focal surgery during the same operation, and this concurrent surgery will put the surgeons in a state of mindful activity and high alertness (Levinthal and Rerup 2006, KC 2014). This, in turn, will enable them to spot the nuances of and differences between the two types of surgery more easily, hence reducing the likelihood of invalid generalizations and negative learning transfer. In addition, the difficulty of performing different operations concurrently leads to enhanced information processing and decision-making ability, which promote efficient learning (Shea and Zimny 1983, Lee and Simon 2004). Indeed, recent medical research has observed that performing multiple surgeries during the same operation can promote more efficient learning for trainee surgeons (Bongers et al. 2015) and that surgeons performing multiple tasks at once are able to successfully reallocate their attention resources (Grierson et al. 2011). That is, performing another task concurrently with the focal one (e.g., performing both CABG and valve replacement concurrently) helps surgeons better comprehend, learn about, and identify the intricacies of the focal task (CABG) itself. We identify two mechanisms of this effect.

First, performing a focal task concurrently with other tasks provides a new context for the focal task. This variation in context will help surgeons develop "implicit learning" (Reber 1989, Wulf and Schmidt 1997). That is, without even realizing it, surgeons will develop critical but highly complex and abstract knowledge about the focal task and its associations (Maskarinec and Thompson 1976). Through this implicit learning process, surgeons will improve their understanding and performance of the focal task. Indeed, researchers have shown that implicit learning promotes neural efficiency (i.e., more expert-like mapping of neural resources in the completion of the task) in surgical training (Zhu et al. 2011). As a result, surgeons become more productive, since they are able to deploy resources more easily to other non-technical aspects of the surgery (Masters et al. 2008, Zhu et al. 2011).

Second, in the acquisition of a cognitive skill, potential errors in initial understanding are gradually detected and eliminated (Fitts 1964). In fact, a widely recognized cognitive theory of learning (ACT-R adaptive control of thought-rational, Anderson 2013) suggests that a *discrimination process* plays a critical role in the learning and acquisition of a cognitive skill (Anderson 1982). This discrimination process helps one narrow and specify the applicability of new procedural knowledge by producing multiple variants on the conditions of the same action. One important benefit of performing a different and concurrent task is therefore that it facilitates this discrimination process. In facing multiple tasks rather than an individual one, the most appropriate course of action for each particular task is better identified and learned by remembering and comparing its variable bindings (Anderson 1982). Considering all the above arguments, we predict that:

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

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

We next focus on the way recent (short-term) non-concurrent variety interacts with long-term non-concurrent variety. We argue that the negative effects of non-concurrent variety on surgeons' decision-making skills will be aggravated by recent non-concurrent exposure to variety.

As discussed in Section 2.3.3, 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, which is also observed among medical staff (Loewenstein and Lerner 2002, Bargh and Williams 2007), 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 similar experiences (Bargh and Williams 2006). The reason is that priming is more likely to occur in the presence or even vestiges of recent relevant behavior (Bargh et al. 2012), since individuals tend to respond unconsciously with their most recent relevant behaviors. We therefore expect that the detrimental effect due to unconscious response (Bargh and Williams 2006) caused by non-concurrent variety will be more significant when a surgeon has recently performed another task independently (non-concurrently).

Exposure to variety can create long-term residual costs on subsequent focal tasks (Allport et al. 1994, Monsell 2003). This long-term residual cost is likely to increase with a recent non-concurrent exposure to variety, which will require further task-set reconfiguration (Rogers and Monsell 1995, Meiran 1996, Monsell 1996, De Jong 2000, Rubinstein et al 2001). That is, after performing a recent but different operation independently, a surgeon will have to modify the set of rules in her cognitive system for performing an operation, and this modification is associated with a switch-and-set-up cost for the individual (Schultz et al. 2003, Staats and Gino 2012). Also, because this task variety will have taken place in a separate operation preceding the focal surgery, it will not be assisted by the adaptive executive control system (Meyer and Kieras 1997, Schumacher et al. 2001) which helps individuals achieve cost-free dual-task performance in switching tasks within the same operation (Schumacher et al. 2001, Hazeltine et al. 2002).

In addition, the negative knowledge-transfer effect (Gick and Holyoak 1987) and invalid generalizations of non-current variety which result in reduced productivity will be further amplified by recent non-concurrent exposure to variety. Because individuals tend to retrieve schemas that they have used recently, even when more plausible and reasonable alternatives exist (Reder 1982), a recent non-concurrent exposure to variety will result in surgeons initially adopting the schema of the recently performed different task, and this schema may not be appropriate for the current task. Also, recent variety may amplify the difference between cognitive perception and the actual level of knowledge, which is detrimental for productivity Hollyoak 1985, Zollo and Reuer 2010). As a result, the probability of negative knowledge transfer due to misjudgments about the similarity between the focal task and other tasks will be higher.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.*

2.3.6 The Moderating Role of Short-Term Concurrent Variety

Finally, we consider the moderating role 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 in the presence of a recent concurrent variety.

As discussed in Section 2.3.4, an important way that concurrent variety improves subsequent focal task productivity is "implicit learning", in which variation in the context helps individuals improve both recall and understanding of the focal task (Maskarinec and Thompson 1976). 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 and promote neural efficiency.

In addition, individuals tend to forget less rapidly as the complexity of the performed task increases (Lance et al. 1998, Nempbhard 2000), while they forget more rapidly when the time interval between two consecutive tasks is increased (Globerson et al. 1989), even in a procedural cognitive task (Nembhard and Uzumeri 2000) such as a cardiac operation. Researchers have also shown that learned skills (both technical and cognitive ones) tend to deteriorate for surgeons after some time and have highlighted the need for periodic remediation of any necessary skills (Kahol et al. 2010). Moreover, it has been suggested that performing multiple surgeries concurrently can increase surgeons' retention of skills (Kahol et al. 2010). Hence we believe that a recent concurrent exposure will significantly reduce the forgetting effect for a surgeon. With reduced forgetting, long-term learning effect from exposure to variety will increase.

Finally, the probability of misguided generalizations will be further decreased when a surgeon recently performs an operation that includes the focal task and another one concurrently. Recent concurrent exposure will allow the surgeon to better understand the differences among different surgery types, which will decrease—if not eliminate—the likelihood of invalid generalizations and the risk of adopting inefficient strategies when subsequently performing the focal task (CABG). 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.*

2.4 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 dataset 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. There are 44 surgeons in our sample: 19 started working after the beginning of our dataset, and 13 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 role 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, different surgery types will result in surgeons performing different set of steps and activities during the surgery, providing many opportunities for learning. However, tasks and activities for the rest of the team members (e.g., the anaesthesiologist preparing the patient, nurses providing the equipment, etc.) are more trivial and very similar. Since our primary research agenda in this paper is to identify the effect of task variety on learning and productivity, we focus on surgeons, who do perform somewhat different activities and tasks. Second, interviews with the medical staff at the hospital confirmed our intuition that we should consider only surgeon members of the team when studying the role of task variety on productivity in this setting.

Our dataset 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 condition prior to operation. Specifically, the hospital labels each patient's case either "severe", "medium" or “mild”. Finally, our dataset includes 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 choose to 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 individual 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 to consider as the focal task for our study.

A surgeon can perform multiple surgery types 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. There are 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 dataset includes 951 complex operations in which a CABG surgery and one of the other types of surgeries were performed concurrently. Our dataset 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 ones; it is worth noting that apart from the 951 complex operations in which a CABG is included, there are 14 more with two surgeries other than CABG performed concurrently).

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

2.4.1 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, in our semiformal interviews, staff members in our hospital also confirmed that lower completion times usually reflect better clinical outcomes.

While our dataset 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, similar to 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). Therefore, in line with our theoretical development, we have decided to keep the scope of our study on productivity performance and use in-hospital mortality as a robustness check to make sure that shorter completion times do not come at the cost of increased mortality rates. One may argue that shorter completion times may also represent inattention to the clinical outcome of the operation. However, we believe that this is not the case in our setting. First, prior research has shown that shorter completion times decrease the probability of post-surgical infections for cardiac surgeries (Gaynes et al. 2001, Gibbons et al. 2011) and can actually improve the patient's clinical outcome (Pisano et al. 2001). Second, we have empirically investigated the association between duration and in-hospital mortality to check whether decreased durations are associated with increased mortality rates. Specifically, we have created four groups of operations according to their duration (i.e., those less than the 25th percentile, between the 25th and 50th percentile, between the 50th and 75th percentile, and higher than 75th percentile) and conducted a chi-square test to examine the relationship between in-hospital mortality and being in one of these four groups. Our results indicate that patient deaths are not evenly spread across these four groups (p-value = 0.023) and that the number of observed in-hospital deaths is significantly higher in the group with longest durations than in other groups. This suggests that shorter completion times are not associated with worse outcomes. This finding is in line with previous medical studies which found that longer surgical durations are associated with higher probabilities of a surgical site infection (SSI) in several surgery types, including cardiac operations (Gaynes et al. 2001, Gibbons et al. 2011).

2.4.1.1 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 2.1 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 2.1 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 of five hours long). 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.

2.4.1.2 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 2.1 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 labeled the patient as such and zero otherwise. We expect "severe" cases to generally last longer than the other two categories.

2.4.1.3 Calculation of Key Independent and Control Variables

Table 2.1 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.

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 (at times t_1 and t_3) 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 3 of the Appendix shows the average, standard deviation, and median for all lead surgeons for these variables.

Table 2.1. 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	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

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 used the Durbin-Watson test, we found 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) explicitly takes into account all of these issues (i.e., xtregar command in Stata).

Also, our analyses of the distribution of dependent and continuous independent variables (*ladder* and *gladder* commands in Stata) revealed skewed distributions. Consequently, and in line with our theoretical arguments and previous studies on conventional learning curve models examining completion times (Argote 1999, Reagans et al 2005), we decided to take the logarithm of all these variables. We also confirmed the normality of residuals and checked for heteroscedasticity by using the

Breusch-Pagan test (1979), which did not reject the null hypothesis, thereby confirming that heteroscedasticity does not pose a threat to our analyses.

Our model is the following:

$$
\begin{aligned}
ln(duration_{ij}) = {} & \beta_0 + \beta_1\, ln(Non\text{-}Concurrent\ Variety_{ij}) \\
& \beta_2\, ln(Concurrent\ Variety_{ij}) + \\
& \beta_3\, ln(Recent\ Non\text{-}Concurrent\ Variety_{ij}) \times ln(Non\text{-}Concurrent\ Variety_{ij}) + \\
& \beta_4\, ln(Recent\ Concurrent\ Variety_{ij}) \times ln(Concurrent\ Variety_{ij}) + \\
& \beta_5\, ln(Recent\ Non\text{-}Concurrent\ Variety_{ij}) + \\
& \beta_6\, ln(Recent\ Concurrent\ Variety_{ij}) + \\
& \beta_7\, ln(Focal\ Experience_{ij}) + \\
& \beta_8\, ln(Team\ Size_{ij}) + \\
& \beta_9\, ln(Individual\ Average\ Experience_{ij}) + \\
& \beta_{10}\, (Severe_{ij}) + \\
& \beta_{11}\, (Mild_{ij}) + \\
& \alpha_i + \\
& t_j + \\
& u_{ij}, \\
& \text{where } u_{ij} = \rho u_{ij-1} + e_{ij}
\end{aligned}
$$

In the above model, α_i represents unobserved lead surgeon fixed effect, and 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 2.2 shows descriptive statistics and correlations among the variables. Table 2.3 shows the results for all our hypotheses. Due to the AR(1) covariance structure we employ, the number of observations in the table is equal to 3,261. In Model 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 2, we add our first variable of interest: *Non-Concurrent Variety*. The adjusted R^2 is increased by 3.5%, and an F-test showed that Model 2 is superior to Model 1 ($p < 0.05$). *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$), providing support for our first hypothesis. In Model 3 we add *Concurrent Variety.* The adjusted R^2 is further increased by 5.41%, and an F-test showed that Model 3 is

2.2 Descriptive Statistics for Lead Surgeons

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
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	5.033	0.971	0	5.829	0.119*	0.579**	1									
4. Recent Concurrent Variety	0.505	0.601	0	1.792	-0.019*	0.494**	0.271**	1								
5. Concurrent Variety	4.920	0.996	0	5.513	0.109**	0.454**	0.592**	0.251**	1							
6. Recent Non-Concurrent Variety x Non-Concurrent Variety	4.834	1.970	0	7.087	0.026*	0.823**	0.605**	0.351**	0.616**	1						
7. Recent Concurrent Variety x Concurrent Variety	4.191	2.004	0	6.977	-0.041*	0.382**	0.455**	0.879**	0.479**	0.550**	1					
8. Focal Experience	5.978	1.106	0	6.477	0.052**	0.354**	0.572**	0.254**	0.695**	0.446**	0.285**	1				
9. Team Size	1.617	0.130	0.693	2.079	0.117**	-0.097*	-0.039**	-0.016	0.065*	-0.300**	-0.281	0.049**	1			
10. Average Individual Experience	6.085	1.100	0	7.701	0.103**	0.101*	0.540**	0.041**	0.601**	0.465**	0.284**	0.686**	0.069**	1		
11. Severe	0.043	0.203	0	1	0.036*	0.035*	0.011*	-0.013	-0.058**	-0.017	-0.009	-0.048**	0.024	-0.064**	1	
12. Mild	0.291	0.454	0	1	-0.061**	-0.011	-0.015+	-0.018	-0.037+	-0.016	0.007	-0.035*	-0.024	-0.047**	-0.136**	1

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

Logged values of all variables except Severe and Mild

Table 2.3 Regression of Task Variety on Surgery Duration only for Lead Surgeons

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.313**	0.341**	0.329**	0.320**
		(0.047)	(0.048)	(0.048)	(0.049)
Concurrent Variety			-0.349**	-0.356**	-0.345**
			(0.050)	(0.050)	(0.051)
Recent Non-Concurrent Variety				0.075**	0.075**
				(0.022)	(0.025)
Recent Concurrent Variety					-0.045*
					(0.022)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.035**	0.037**
				(0.011)	(0.013)
Concurrent Variety x Recent Concurrent Variety					-0.024*
					(0.010)
Focal Experience	-0.127**	-0.130**	-0.122**	-0.119**	-0.119**
	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Team Size	0.109**	0.151**	0.137**	0.129**	0.128**
	(0.028)	(0.036)	(0.036)	(0.036)	(0.036)
Average Individual Experience	-0.031**	-0.028**	-0.023*	-0.024*	-0.024*
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Severe	0.025*	0.024*	0.023*	0.022*	0.023*
	(0.012)	(0.011)	(0.011)	(0.010)	(0.011)
Mild	-0.013*	-0.013*	-0.012*	-0.011+	-0.011+
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	3.810**	3.743**	3.845**	3.921**	3.929**
	(0.110)	(0.116)	(0.116)	(0.118)	(0.119)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.143	0.148	0.156	0.159	0.160
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
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

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

superior to Model 2 ($p < 0.01$). *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$), supporting Hypothesis 2.

In Model 4 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 1.92%, and an F-test indicated that Model 4 is superior to Model 3 ($p < 0.01$). We also see that *Non-Concurrent Variety x Recent Non-Concurrent Variety* is significant ($p < 0.01$) and positive, providing support for our third Hypothesis. Finally, in Model 5 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 0.63%, and an F-test indicated that Model 5 is superior to Model 4 ($p < 0.05$). *Concurrent Variety x Recent Concurrent Variety* is significant ($p < 0.05$) and negative, providing support for Hypothesis 4.

According to Model 3, an increase of 20% in *Non-Concurrent Variety* increases duration by 6.82% (20.31 minutes), whereas such an increase of *Concurrent Variety* decreases duration by 6.98% (20.78 minutes). Next, we repeat our analysis using all the surgeons of the team (not just the lead surgeon). Specifically, we include all surgeons of the team when calculating our independent variables and repeat our analysis while controlling for the experience of the other members (in this case, we calculate *Non-Concurrent Variety*, *Concurrent Variety,* and *Focal Experience* using all the surgeons in the team and then taking the average and *Individual Average Experience* using all team members excluding surgeons). Table 2.4 shows descriptive statistics and correlations among the variables. Table 2.5 shows the results. In Model 1 we include only our control variables. In Model 2, we add our first variable of interest, *Non-Concurrent Variety*. The adjusted R^2 is increased by 12.59%, and an F-test showed that Model 2 is superior to Model 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 we add *Concurrent Variety.* The adjusted R^2 is further increased by 8.70%, and an F-test showed that Model 3 is superior to Model 2 (p<0.01). *Concurrent Variety* has a negative and significant coefficient ($p<0.01$), supporting Hypothesis 2.

In Model 4 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 0.57%, and an F-test indicated that Model 4 is superior to Model 3 (p<0.05). We also see that *Non-Concurrent Variety x Recent Non-Concurrent Variety* is significant and positive (p<0.01), providing support for our third hypothesis. Finally, in Model 5 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 1.70%, and an F-test indicated that Model 5 is superior to Model 4 (p<0.01). *Concurrent Variety x Recent Concurrent Variety* is significant and negative (p<0.01), providing support for Hypothesis 4. One limitation of our fixed-effects AR(1) regression based on Baltagi and Wu (1999) is that it was not possible to report robust standard errors. The Breusch-Pagan test (1979) revealed 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 2.6 and 2.7 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. We next investigate the economic significance of our main variables of interest. According to Model 3, an increase of 20% in *Non-Concurrent Variety* increases duration by 4.76% (14.17 minutes), whereas such an increase of *Concurrent Variety* decreases duration by 4.06% (12.09 minutes). These findings suggest that not only are the effects of concurrent and non-concurrent variety statistically significant, but their practical effects on surgery completion times are also considerable. Next, we conduct post hoc plots for Hypotheses 3 and 4 using the lead

2.4 Descriptive Statistics for all Surgeons

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1. Duration	5.681	0.176	4.094	6.579	1											
2. Recent Non-Concurrent Variety	1.038	0.573	0	1.792	0.017**	1										
3. Non-Concurrent Variety	5.129	1.356	0	7.486	0.109**	0.356**	1									
4. Recent Concurrent Variety	0.831	0.657	0	1.609	-0.045**	0.420**	0.191**	1								
5. Concurrent Variety	4.839	1.276	0	6.912	0.078**	0.373**	0.692**	0.217**	1							
6. Recent Non-Concurrent Variety x Non-Concurrent Variety	5.603	3.474	0	11.379	0.056**	0.818**	0.634**	0.640**	0.380**	1						
7. Recent Concurrent Variety x Concurrent Variety	4.205	3.572	0	9.677	-0.033*	0.426**	0.407**	0.429**	0.847**	0.483**	1					
8. Focal Experience	6.089	1.214	0	8.035	0.082**	0.357**	0.690**	0.204**	0.692**	0.629**	0.419**	1				
9. Team Size	1.617	0.130	0.693	2.079	0.117**	0.053**	0.067**	0.007	0.039*	0.085**	0.019	0.052**	1			
10. Average Individual Experience	5.662	1.036	0	7.312	0.110**	0.188**	0.650**	0.080**	0.640**	0.437**	0.271**	0.651**	0.071**	1		
11. Severe	0.043	0.203	0	1	0.036*	-0.029+	-0.064**	-0.010	-0.070**	-0.048**	-0.023	-0.064**	0.024	-0.048**	1	
12. Mild	0.291	0.454	0	1	-0.061**	0.011	-0.041*	0.026	-0.032+	-0.003	0.016	-0.038*	-0.024	-0.050**	-0.136**	1

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

Logged values of all variables except Severe and Mild

Table 2.5 Regression of Task Variety on Surgery Duration for all Surgeons

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.171**	0.238**	0.269**	0.227**
		(0.029)	(0.029)	(0.022)	(0.030)
Concurrent Variety			-0.203**	-0.180**	-0.261**
			(0.035)	(0.035)	(0.036)
Recent Non-Concurrent Variety				0.029	0.053*
				(0.021)	(0.022)
Recent Concurrent Variety					-0.044*
					(0.019)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.013**	0.009*
				(0.004)	(0.004)
Concurrent Variety x Recent Concurrent Variety					-0.020**
					(0.004)
Focal Experience	-0.187**	-0.186**	-0.193**	-0.198**	-0.195**
	(0.017)	(0.019)	(0.022)	(0.022)	(0.023)
Team Size	0.067*	0.071**	0.070**	0.079**	0.081**
	(0.027)	(0.025)	(0.024)	(0.025)	(0.025)
Average Individual Experience	-0.031**	-0.028**	-0.029**	-0.029**	-0.029**
	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Severe	0.032*	0.031*	0.034*	0.035*	0.032*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Mild	-0.013*	-0.012+	-0.010	-0.010	-0.010
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	4.110**	4.734**	4.895**	4.807**	4.801**
	(0.110)	(0.114)	(0.113)	(0.114)	(0.114)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.143	0.161	0.175	0.176	0.179
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
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

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

surgeon and then all surgeons of our teams to examine the moderating effects described by Aiken and West (1991) and Dawson and Richter (2006). We divide our sample into a subset with values above the median for Recent Non-Concurrent Variety and a subset with values below the median for Recent Non-Concurrent Variety and plot Non-Concurrent Variety and Duration (Figures 2.1 and 2.2). Notice that we employ logged variables in the figures. We also divide the sample into a subset with values above the median for Recent Concurrent Variety and one with values below Recent Concurrent Variety and plot Concurrent Variety and Duration (Figures 2.3 and 2.4). In dividing our sample, we decided to use the median instead of the mean plus one standard deviation and the mean minus one standard deviation because in the latter case our sample was dramatically decreased. Although all plots reveal the moderating effects as proposed in Hypotheses 3 and 4, the effect of the moderation is quite small in terms of economic significance (also note the coefficients of the interaction terms in our models). That is, combined with the main effects of concurrent and non-concurrent task variety, and their short term effects, we observe quite modest moderation effects in terms of practical magnitude in our sample. This is not surprising in our surgery setting, since the primary drivers of any surgery completion time are, first and foremost, clinical factors. Consequently, we believe that our theoretical insights on moderation effects are useful despite their limited practical applicability in our setting with their small effect sizes. In addition, in other professional service settings where external factors (e.g., the patient's clinical condition) are not as dominant in driving productivity outcomes (such as legal services, consulting, etc.), we expect such moderation effects to be sizeable and of economic significance.

Figure 2.1 Interaction Plot for H3 for Lead Surgeons

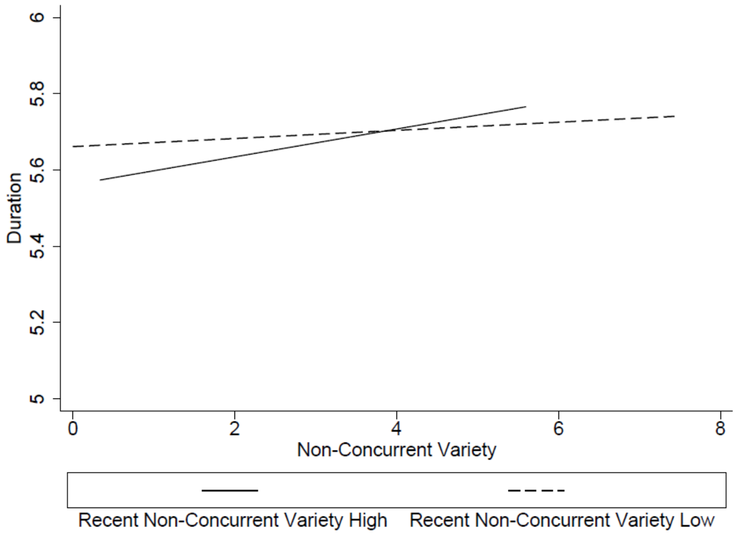


Figure 2.2 Interaction Plot for H3 for all Surgeons

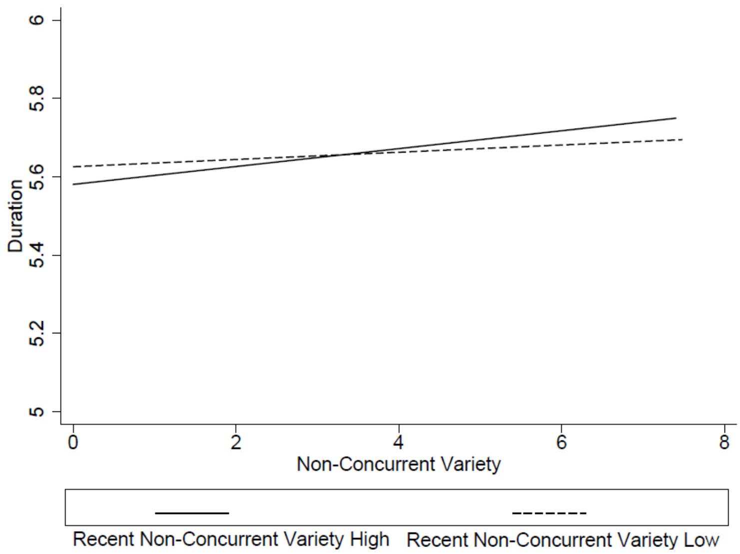


Figure 2.3 Interaction Plot for H4 for Lead Surgeons

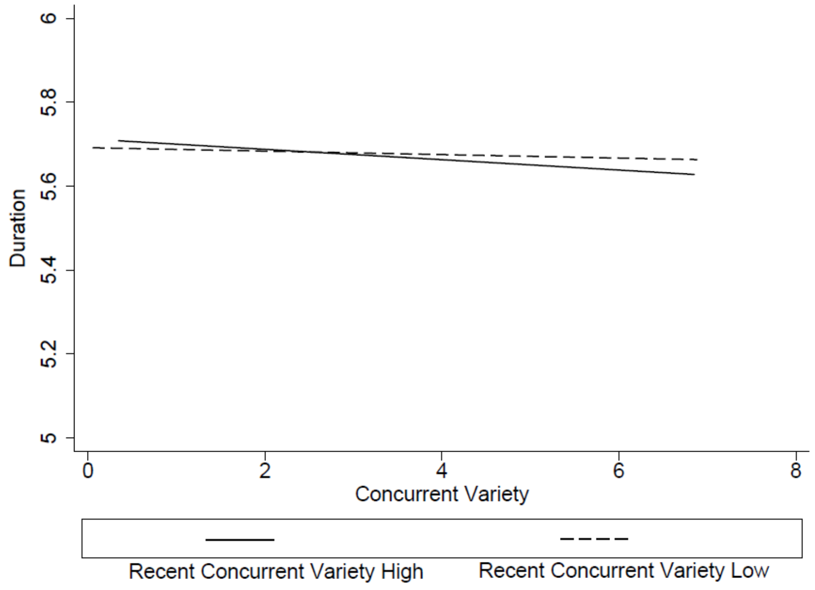
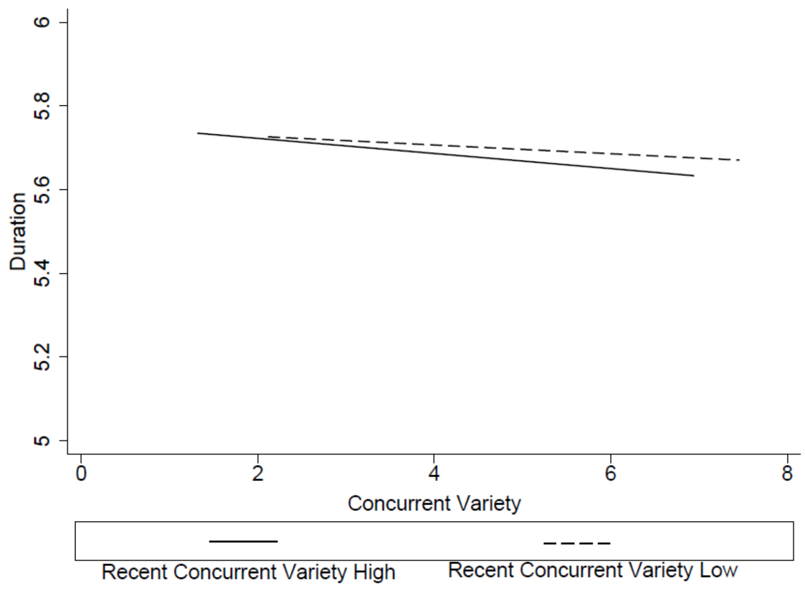


Figure 2.4 Interaction Plot for H4 for all Surgeons



2.6 Robustness Checks

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

2.6.1 Data and Variables' Operationalization

As in any empirical study, our dataset is limited. A potential concern is that we have no data for surgeries or surgeons prior to the beginning of our dataset, which may influence our results. To address this, we repeat our analysis after excluding different time intervals from the beginning of our dataset. 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 qualitatively. We therefore believe that missing experience (i.e., missing data prior to the beginning of our dataset) 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 for 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 changes as expected, our primary findings remain the same.

We also replaced 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 dataset. Our results remain the same.

Finally, one may be concerned that the effects we observe may only work on single focal operations (CABG-only) and may not hold on 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.

2.6.2 Surgery Assignments

An additional concern could be the possibility that more severe cases might be assigned to more experienced lead surgeons or, similarly, easier cases to less experienced lead surgeons. To deal with this, 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 2.8 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 qualitatively. We also investigated the spread of deaths among surgeons and examined any potential correlations of these deaths with our key variables. We find that these in-hospital deaths are spread quite evenly across surgeons and show no correlation with our key variables other than the “severe” patient condition.

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 examined this issue: We repeated our analysis after excluding the severe cases (using only the mild and medium ones). Then we repeated our analysis after excluding only the mild cases from our initial sample (using only the medium and severe ones). Finally, we also repeated our analysis using only the medium cases. In all three cases, while the magnitude of effects changed considerably as expected, our results remained the same.

2.6.3 Instrumental Variable Approach

We conducted fixed-effects regressions, which control for all observed and unobserved time-invariant heterogeneity across lead surgeons through a de-meaning process, as well as the further analyses of surgery assignments outlined above. However, these analyses might not have fully addressed potential time-varying individual effects which are unobservable to us but may affect lead surgeons’ task variety and productivity at the same time.

To address this, we use an instrumental variables approach with 2SLS specification. Specifically, we use the number of public holidays of the country in which the hospital is located and the number of the lead surgeon's vacation days as instruments for surgeons' concurrent and non-concurrent variety. This approach is similar to Lambrecht et al. (2011) which uses vacations and public holidays as instruments for interruption of individuals' adoption process of online banking service. We use variation in time intervals of public holidays and individuals' vacation days in our two-stage estimation strategy. That is, while a lead surgeon's exposure to variety is influenced (i.e., interrupted) during holidays and vacations (since no operation will take place in these time intervals), the number of days spent in these intervals is unrelated to any CABG assignment and selection issues in the hospital. Also, the cumulative number of such days off should not have any influence on surgery durations. Consequently, we believe that our instruments satisfy the exclusion restriction.

We create two variables for our instruments: *Public Holidays* and *Days of Vacation*. For every operation, we calculate the number of public holidays prior to that operation and take the sum. Finally, since exposure to variety increases as time goes by, when we define our variable *Public Holidays,* we use the number of days that each lead surgeon has been working at the hospital minus the sum of the public holidays up to this point. We expect to observe a positive association between this variable and lead surgeons' exposure to variety.

Regarding *Days of Vacations*, we define vacations if the lead surgeon does not appear in our sample for two weeks or more (we also considered 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. So, for every operation, we calculate total days of vacation the lead surgeon has taken up to the current operation. Finally, since exposure to variety increases as time goes, when we define *Days of Vacations,* we use the number of days that each lead surgeon has been working at the hospital minus the number of days she has taken as vacation up to this point. We expect to observe a positive effect of this variable on lead surgeons' exposure to variety. Tables 2.9, 2.10, and 2.11 show the results of our IV approach.

In Table 2.9, we only consider *Non-Concurrent Variety* and use *Public Holidays* as the instrument. We first evaluate the quality of *Public Holidays* as the instrument. Our first-stage estimation in Model 1 reveals a significant positive effect ($p < 0.01$) of *Public Holidays* on *Non-Concurrent Variety* with an F-statistic well above the common threshold of 10, indicating that it is not a weak instrument. Our second-stage estimation, Model 2 in Table 2.9, provides support for our first hypothesis: *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$).

In Table 2.10, we focus only on *Concurrent Variety* and use *Public Holidays* as the instrument. We examine the first stage estimation with *Public Holidays* as the instrumental variable and observe a significant ($p < 0.05$) positive effect of *Public Holidays* on *Concurrent Variety*. The F-statistic is also well above the common threshold of 10, indicating that it is not a weak instrument. Our second-stage results in Model 2 of Table 2.10 provide support for our second hypothesis: *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$).

In Table 2.11, we use both *Public Holidays* and *Days of Vacations* as instruments to *Non-Concurrent Variety* and *Concurrent Variety* in two stage estimation. As before, we first evaluate the quality of our instruments with first-stage model and F-statistic (again well above 10), and confirmed the strength of the instruments. The second stage estimation, Model 3 from Table 2.11, provides support for both H1 and H2. Specifically, *Non-Concurrent Variety* has a positive and significant coefficient ($p < 0.01$) providing support for our first hypothesis, and *Concurrent Variety* has a negative and significant coefficient ($p < 0.01$), supporting Hypothesis 2. We are therefore confident that endogeneity does not bias our results.

Table 2.9 Regression of Task Variety with Instrumental Variable Approach

Variable	Non-Concurrent Variety	Duration
	Model: (1)	(2)
Non-Concurrent Variety		0.192**
		(0.017)
Public Holidays	0.003**	
	(0.001)	
Focal Experience	0.031**	-0.131**
	(0.019)	(0.038)
Team Size	-3.200**	0.166**
	(0.067)	(0.051)
Average Individual Experience	0.071*	-0.078**
	(0.028)	(0.011)
Severe	0.108**	0.028*
	(0.034)	(0.012)
Mild	-0.005	-0.011
	(0.015)	(0.010)
Constant	5.573**	4.252**
	(0.300)	(0.913)
Observations (N)	3273	3261
Adjusted R^2	0.551	0.139
Day of the Week Fixed Effect	Yes	Yes
Month Fixed Effect	Yes	Yes
Year Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes

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

Table 2.10 Regression of Task Variety with Instrumental Variable Approach

Variable	Concurrent Variety	Duration
	Model: (1)	(2)
Concurrent Variety		-0.118**
		(0.026)
Public Holidays	0.002*	
	(0.001)	
Focal Experience	0.030**	-0.233**
	(0.018)	(0.061)
Team Size	-3.029**	0.145**
	(0.064)	(0.042)
Average Individual Experience	0.044	-0.071**
	(0.027)	(0.013)
Severe	0.097**	0.029*
	(0.032)	(0.013)
Mild	-0.002+	-0.012
	(0.001)	(0.018)
Constant	4.990**	5.900**
	(0.284)	(1.373)
Observations (N)	3273	3261
Adjusted R^2	0.547	0.132
Day of the Week Fixed Effect	Yes	Yes
Month Fixed Effect	Yes	Yes
Year Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes

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

Table 2.11 Regression of Task Variety with Instrumental Variable Approach

Variable	Non-Concurrent Variety	Concurrent Variety	Duration
	Model: (1)	(2)	(3)
Non-Concurrent Variety			0.205**
			(0.016)
Concurrent Variety			-0.124**
			(0.011)
Public Holidays	0.003**	0.002**	
	(0.001)	(0.001)	
Days of Vacations	0.003**	0.002**	
	(0.001)	(0.001)	
Focal Experience	0.042**	0.041**	-0.103**
	(0.018)	(0.017)	(0.031)
Team Size	-3.200**	-3.033**	0.159**
	(0.067)	(0.063)	(0.44)
Average Individual Experience	0.075**	0.050+	-0.078**
	(0.027)	(0.026)	(0.015)
Severe	0.109**	0.099**	0.027*
	(0.034)	(0.032)	(0.012)
Mild	-0.004	-0.001	-0.007
	(0.015)	(0.014)	(0.008)
Constant	5.296**	4.810**	5.719**
	(0.156)	(0.147)	(0.700)
Observations (N)	3273	3273	3261
Adjusted R^2	0.554	0.549	0.149
Day of the Week Fixed Effect	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes

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

2.7 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).

In this study, we examine how exposure to task variety in different forms may influence surgeons' productivity over time. We introduce time as a new dimension in the way an individual may perform a variety of tasks, and we show that timing can alter the effect of task variety on individual learning and subsequent productivity. More precisely, 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 vs. in 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 seeks 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, allows us to investigate the influence of exposure to variety on individual learning and productivity over time. In addition, because we are 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 by exogenous (i.e., medical) factors, our results are not likely to be affected by endogeneity concerns.

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 4.06% (which, in the case of our surgeons, translates to 18 additional CABG operations per year). On the other hand, we observe that performing other tasks in isolation (non-concurrently) does not provide such learning and productivity improvement opportunities. In contrast, such non-concurrent variety appears to be detrimental to productivity. Our results suggest that a 20% increase in non-concurrent exposure to variety can decrease productivity by 4.76% (which, for our surgeons, translates to around 21 additional CABG operations per year). To maximize individual learning and improve productivity in common tasks, therefore, a worker may consider pairing her 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.

As in all empirical studies, our findings and conclusions are subject to limitations. First, our dataset includes limited information about the patients' condition before the operation. Specifically, the hospital labels each patient's case as mild, medium, or severe. Ideally, we would like to have more detailed information about patients' condition, such as their EuroScore or Higgins score, but this was not available in our archival data. In addition, our dataset comes from a single hospital. One should therefore be careful when interpreting the results of our study. In addition, despite the fact that cardiac surgery is an appealing setting in which to study the effect of task variety on learning and productivity, generalizing our results to other professional service settings requires a cautious approach, since specific task dynamics and contextual factors affecting learning mechanisms may be different in other settings.

2.8 Appendix

Table 2.6 Regression of Task Variety on Surgery Duration only for Lead Surgeons with robust standard errors

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.323**	0.356**	0.337**	0.328**
		(0.054)	(0.054)	(0.055)	(0.058)
Concurrent Variety			-0.366**	-0.373**	-0.362**
			(0.056)	(0.057)	(0.061)
Recent Non-Concurrent Variety				0.066**	0.065*
				(0.021)	(0.029)
Recent Concurrent Variety					-0.049+
					(0.028)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.032**	0.033**
				(0.011)	(0.013)
Concurrent Variety x Recent Concurrent Variety					-0.024*
					(0.011)
Focal Experience	-0.128**	-0.133**	-0.125**	-0.122**	-0.122**
	(0.016)	(0.018)	(0.018)	(0.019)	(0.019)
Team Size	0.101**	0.148**	0.134**	0.127**	0.126**
	(0.027)	(0.039)	(0.039)	(0.038)	(0.038)
Average Individual Experience	-0.033**	-0.027**	-0.027*	-0.023*	-0.023*
	(0.007)	(0.010)	(0.013)	(0.011)	(0.011)
Severe	0.031**	0.029*	0.027*	0.028*	0.028*
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Mild	-0.011+	-0.011+	-0.010	-0.010	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	3.786**	3.713**	3.818**	3.877**	3.882**
	(0.116)	(0.127)	(0.127)	(0.127)	(0.127)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.142	0.147	0.161	0.162	0.163
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
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

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

Table 2.7 Regression of Task Variety on Surgery Duration only for all Surgeons with robust standard errors

Variable	Duration				
	Model: (1)	(2)	(3)	(4)	(5)
Non-Concurrent Variety		0.166**	0.233**	0.262**	0.221**
		(0.032)	(0.033)	(0.034)	(0.031)
Concurrent Variety			-0.199**	-0.184**	-0.271**
			(0.036)	(0.036)	(0.036)
Recent Non-Concurrent Variety				0.053*	0.050*
				(0.024)	(0.022)
Recent Concurrent Variety					-0.055*
					(0.024)
Non-Concurrent Variety x Recent Non-Concurrent Variety				0.013**	0.009*
				(0.004)	(0.004)
Concurrent Variety x Recent Concurrent Variety					-0.022**
					(0.004)
Focal Experience	-0.181**	-0.190**	-0.199**	-0.192**	-0.197**
	(0.017)	(0.020)	(0.024)	(0.024)	(0.024)
Team Size	0.069*	0.074**	0.075**	0.085**	0.087**
	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)
Average Individual Experience	-0.031**	-0.021*	-0.020*	-0.023*	-0.021*
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Severe	0.031*	0.033*	0.030*	0.031*	0.031*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Mild	-0.011+	-0.011+	-0.009	-0.009	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	4.117**	4.631**	4.695**	4.709**	4.703**
	(0.118)	(0.125)	(0.120)	(0.121)	(0.120)
Observations (N)	3261	3261	3261	3261	3261
Adjusted R^2	0.146	0.166	0.180	0.181	0.184
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
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes

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

Table 2.8 Distribution of Cases among Lead Surgeons

Patient's Group	Pearson Chi (11)	P-value
Mild	10.900	0.452
Medium	11.515	0.401
Severe	14.356	0.214

2.9 References

Aiken, L. S., S. G. West. 1991. *Multiple regression: Testing and interpreting interactions*. Sage, Thousand Oaks, CA.

Allport D. A., G. Wylie. 1999. Task-switching: Positive and negative priming of task-set. G.W. Humphreys, J. Duncan, A.M. Treisman eds. *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, MA, 421–452.

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

Anderson, J. R. 1982. Acquisition of cognitive skill. *Psych. Rev*. 89(4) 369.

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

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

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

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

Balch, C.M., T. Shanafelt. 2011. Combating stress and burnout in surgical practice: A review. *Thoracic Surgical Clinics*. 21(3) 417-430.

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 Psych. Sci.* 15(1) 1-4.

Bargh, J. A., E. L. Williams. 2007. On the automatic or nonconscious regulation of emotions. J. Gross eds. *Handbook of Emotion Regulation*. Guilford Press, NY, 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.

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

Bongers, P. J., D. P. van Hove, L. P. Stassen, J. Dankelman, H. W. Schreuder. 2015. A new virtual-reality training module for laparoscopic surgical skills and equipment handling: Can multitasking be trained? A randomized controlled trial. *J. of Surgical Education.* 72(2) 184-191.

Breusch, T. S., A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: J. of the Econom. Soc.* 47 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 Sci.* 58(4) 708-722.

Cohen, M. D., P. Bacdayan. 1994. Organizational routines are stored as procedural memory: Evidence from a laboratory study. *Organ. Sci.* 5(4) 554-568.

Cuschieri, A. N. Francis, J. Crosby, G. B. Hanna. 2001. What do master surgeons think of surgical competence and revalidation? *Amer. J. of Surgery* 182 110-116.

Dawson, J. F., A. W. Richter. 2006. Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. *J. of Appl. Psych.* 91(4) 917-926.

De Jong, R. 2000. An intention-activation account of residual switch costs. S. Monsell. J. Driver eds. *Control of Cognitive Processes: Attention and Performance XVIII.* MIT Press, Cambridge, MA. 357-376.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Admin. Sci. Quart.* 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*. 34(2) 197-224.

Fitts, P. M. 1964. Perceptual-Motor Skill Learning. A. W. Melton. *Categories of Human Learning*. Academic Press, New York, NY.

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 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.

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. Admin. Sci. Quart. 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. *J. of Experiment. Psych.* 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*, 16 661–673.

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 eds. *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, MA.

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

Haleblian, J., S. Finkelstein. 1999. The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective. *Admin. Sci. Quart.* 44(1) 29-56.

Hazeltine, E., D. Teague, R. B. Ivry. 2002. Simultaneous dual-task performance reveals parallel response selection after practice. *J. of Exper. Psych. Human Perception and Performance*. 28(3) 527.

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

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

Holland, J. H. 1986. Escaping brittleness: The possibilities of general purpose learning algorithms applied to parallel rule-based system. R. S. Michalski, J. Carbonell, T. Mitchell eds. *Machine Learning II.* Kauffman, Los Altos, CA, 593-623.

Holyoak, K. J. 1985. The pragmatics of analogical transfer. G. H. Bower eds. *The Psychology of Learning and Motivation*, Vol. 19. Academic Press, New York, 59-87.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Production and Oper. 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. Sci.* 53(1): 61–77.

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

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

Kahol, K., A. Ashby, M. Smith, J. Ferrara. 2010. Quantitative evaluation of retention of surgical skills in simulation. *J. 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 Oper. Management.* 14(4) 618-633.

KC, D. S., 2014. Does Multitasking Improve Performance? Evidence from the Emergency Department. *Manufacturing & Service Oper. Management.* 16(2) 168–183.

Lambrecht, A., Seim K.,Tucker C. 2011. Stuck in the adoption funnel: The effect of interruptions in the adoption process on usage. *Marketing Sci.* 30(2) 355-367.

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, MA.

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

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

Lee, T. D., D. Simon. 2004. Contextual Interference. Williams, A. M., N. J. Hodgeseds eds. *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. *Psych. Science.* 15 337-341.

Levinthal, D. Rerup C. 2006. Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organ. Sci.* 17(4) 502–513.

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

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, NY. 619-642.

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

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* 4 741–746.

Masters, R. S. W., C. Y. Lo, J. P. Maxwell, N. G. Patil. 2008. Implicit motor learning in surgery: implications for multi-tasking. *Surgery.* 143(1) 140-145.

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

Meiran, N. 1996. Reconfiguration of processing mode prior to task performance. *J. of Experimen. Psych.: Learn., Memory, and Cognition.* 22(6) 1423-1442.

Meyer, D. E., D. E. Kieras. 1997. A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psych. Rev.* 104(1) 3.

Monsell, S. 1996. Control of mental processes. V. Bruce. *Unsolved mysteries of the mind: Tutorial essays in cognition.* Psychology Press. 93-148.

Monsell, S. 2003. Task switching. *Trends in Cognitive Sci*. 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*. 55(11) 1861-1876.

Nembhard, D. A. 2000. The effect of task complexity on learning and forgetting: A field study. *Human Factors.* 42(2) 272-286.

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

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. *J. of Organ. Behav*. 27(7) 941-966.

Orri, M., A. Revah-Lévy, O. Farges. 2015. Surgeons' emotional experience of their everyday practice - a qualitative study. *PLoS One*. 10(11).

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*. 47(6) 752-768.

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

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*. 51(6) 869-881.

Reber, A. 1989. Implicit learning and tacit knowledge. *J. Experiment. Psych.* 118 219–235.

Reder, L. M. 1982. Plausibility judgements versus fact retrieval: Alternative strategies for sentence verification. *Psych. Rev.* 19 90-137.

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

Rogers, R. D., S. Monsell. 1995. Costs of a predictable switch between simple cognitive tasks. *J. of Experiment. Psych.: General*. 124(2) 207.

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

Rubinstein, J. S., D. E. Meyer, J. E. Evans. 2001. Executive control of cognitive processes in task switching. *J. of Experiment. Psych.: Human Perception and Performance*. 27(4) 763.

Sarason I.G., G.R. Pierce. 1996. *Cognitive Interference*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schaverien, M. V. 2010. Development of expertise in surgical training. *J. of Surgical Education.* 67(1) 37-43.

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

Schultz K. L., J. O. McClain, L. J. Thomas. 2003. Overcoming the dark side of worker flexibility. *J. Oper. Management* 21(1) 81–92.

Schumacher, E. H., T. L. Seymour, J. M. Glass, D. E Fencsik, E. J. Lauber, D. E. Kieras, D. E. Meyer. 2001. Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psych. Sci.* 12(2) 101-108.

Shea, J. B., S. T., Zimny. 1983. Context effects in memory and learning movement information. R. H. Magil eds. *Memory and Control of Action*. North Holland, Netherlands, 345-366.

Skaugset, L.M., S. Farrell, M. Carney, M. Wolff, S. A. Santen, M. Perry, S. J. Cico. 2015. Can you multitask? Evidence and limitations of task switching and multitasking in emergency medicine. *Ann. Of Emergency Medicine*. Articles in Press.

Spencer, F.C. 1978. Teaching and measuring surgical techniques-the technical evaluation of competence*. Bull. of the Amer. College of Surgeons* 63 9-12.

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

Thurstone, L. L. 1919. The learning curve equation. *Psych. Monographs: General and Appl.* 26(3) i-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*. 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*. 53(6) 894-907.

Tversky, A. 1977. Features of similarity. *Psych. Rev.* 84(4) 327.

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

Waszak, F., B. Hommel, A. Allport. 2003. Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psych.* 46(4) 361-413.

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 Amer. J. of Surgery*. 191(1) 5-10.

Wulf, G., R. A. Schmidt. 1997. Variability of practice and implicit motor learning. *J. Experiment. Psych.: Learn. Memory, and Cognition*. 23 987–1006.

Wylie, G., A. Allport. 2000. Task switching and the measurement of “switch costs.” *Psych. Res.* 63(3–4) 212–233.

Youngson, G. G. 2000. Surgical competence: Acquisition, measurement and retention. *Edinburgh: Royal College of Surgeon of Edinburgh*.

Zhu, F. F., J. M. Poolton, M. R. Wilson, Y. Hu, J. P. Maxwell, R. S. W. Masters. 2011. Implicit motor learning promotes neural efficiency during laparoscopy. *Surgical Endoscopy.* 25(9) 2950-2955.

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

Chapter three – Team Familiarity and Productivity in Cardiac Surgery Operations: The Effect of Dispersion, Bottleneck and Task Complexity

Fluid teams are commonly used by a variety of organizations to perform similar and repetitive yet highly critical and knowledge-intensive tasks. In this study, we develop and test a model of knowledge transfer based on team composition dynamics in fluid team operations. Using a granular dataset of 6,206 cardiac surgeries from the cardiac unit of a private hospital in Europe over more than seven years, our study offers a micro-founded account of how team familiarity (i.e., shared work experience) influences team productivity. We propose that in addition to average team familiarity, managers should also consider more nuanced team composition dynamics including familiarity dispersion, bottlenecks, how familiarity is gained, and what kind of tasks benefit the most from familiarity. We observe that teams with high dispersion of pairwise familiarity exhibit lower team productivity, and the existence of a "bottleneck-pair" may significantly hinder overall knowledge transfer capability, thus, productivity of fluid teams. In addition, we observe that the higher the percentage of familiarity gained from complex tasks, the higher the productivity of the team. Finally, our results suggest that the positive effect of average team familiarity on productivity is enhanced when performing more complicated tasks. Our study provides new operational insights to improve productivity of fluid teams with better team composition strategies.

3.1. Introduction

The modus operandi of many of today's organizations in knowledge-intensive environments is to employ fluid teams (Bushe and Chu 2011, Huckman et al. 2009, Edmondson and Nembhard 2009). One can observe fluid teams in a variety of settings, from scientific research collaborations and management consulting teams to surgical teams, flight crews, and product development teams. Fluid teams have a number of unique characteristics: They have no permanent memberships, they operate as a team only for a limited time or for a specific task, and they normally have clear hierarchies, roles, and task responsibilities. After their job is over, these teams dissolve, but many members may work again with each other as part of another team (Cohen and Bailey 1997, Paletz and Schunn 2011, Summers et al. 2012).

A key management challenge that is particularly pertinent to fluid team operations is team composition. If individuals operate as part of a team only temporarily, how should they be assigned to various teams to make sure that these teams can perform their tasks more efficiently and hence achieve higher productivity over time? Our study seeks to address this question by focusing on a major element of team composition: team familiarity (i.e., shared work experience among team members).

Previous research has identified team familiarity as an important performance driver for teams with a primary focus on the role of average team familiarity (e.g., average of pairs' familiarity) and reported mostly beneficial effects (Edmondson 1999, Faraj and Sproull 2000, Reagans et al. 2005, Espinosa et al. 2007, Huckman et al. 2009, Huckman and Staats 2011) with the exception of two studies which observed diminishing returns in the long run (Katz 1982, Berman et al. 2002).

Despite these studies, several important questions regarding the effect of team familiarity on team productivity have remained unaddressed. First, previous research has mainly focused on the role of the first moment of familiarity distribution among team members (i.e., average familiarity) and has not accounted for two important distributional characteristics that go beyond team averages: (i) dispersion of familiarity among team members (i.e., the amount of variation in pairs' past collaborations, second moment of familiarity distribution), and (ii) existence of a bottleneck pair in the team (e.g., somewhat similar to the third moment which captures how skewed the lowest familiar pair is). Fluid teams are composed of members who have various levels of shared work experience in different pairs, and hence their familiarity distributions might be substantially different despite having similar average familiarity levels. Recent research suggests that key properties of a distribution such as dispersion constitutes an inherent property of every team-level mechanism and hence a significant factor in understanding the effect of that mechanism on team productivity (De Jong and Dirks 2012). As a result, any model that solely focuses on team averages and neglects important distributional properties can lead to biased and ambiguous results for the examined phenomenon (Cole et al. 2011) as it is highly likely to be underspecified (Dineen et al. 2007). Consequently, we first examine how team familiarity dispersion and the existence of bottleneck pairs might influence productivity of fluid teams. Next, inspired by the concept of "bottleneck" in traditional operations settings which limits a system's capacity; we define and identify an analogous construct "bottleneck-pair" for fluid teams. We argue that the existence of a pair with very low shared experience-compared to the average familiarity of the team, may act as a bottleneck for the team. It is worth noting that although the most famous example of a bottleneck in the operations management literature is a person (i.e., Herbie in the hiking track) (Goldratt and Cox 1984), there is little work that relates system bottleneck with persons and teams (Boudreau et al. 2003).

Second, although previous studies have examined various contingent factors on the effect of team familiarity on team performance, such as geographic dispersion and team size (Espinosa et al. 2007), interpersonal diversity in terms of customer experience (Huckman and Staats 2011), and hierarchical roles (Staats 2012), this line of research has not examined the role of task type (i.e., complexity) on the relationship between team familiarity and productivity. On this front, we examine two important questions. First, does familiarity gained from a higher ratio of complex tasks result in higher productivity? Second, how does complexity of the focal task moderate the relationship between team familiarity and productivity? These questions are important from a practical viewpoint too, because fluid teams essentially perform related but not identical jobs over time which may significantly differ in terms of complexity (e.g., consider flight crews or surgical teams where each surgery or flight is different in complexity depending on patient or weather conditions).

Overall, in this study, we focus on four questions related to familiarity in fluid teams. We first look at the distribution of familiarity among team members by examining (i) the effect of team familiarity dispersion, and (ii) the existence of a bottleneck pair on productivity. We then focus on task complexity and investigate (iii) whether familiarity gained from a higher percentage of complex tasks in the past provide higher productivity benefits, and (iv) whether a more complex focal task is more conducive for the positive effects of team familiarity on productivity.

We test our hypotheses using a unique dataset of 6,206 cardiac surgeries from the cardiac unit of a private hospital in Europe over 87 months. With clear team member roles and responsibilities, limited team durations, and repeated interactions among individuals as part of different teams, our setting is ideal to investigate fluid teams (Edmondson et al. 2001). In addition, the presence of similar and repetitive, yet highly critical and complex tasks make cardiac operations a very suitable context to examine the role of team composition on productivity. Our results indicate that teams with high dispersion of pairwise familiarity exhibit lower team productivity (i.e., higher operation completion times), and the existence of a "bottleneck-pair" may significantly reduce productivity. In addition, we observe that the higher the percentage of familiarity gained from complex tasks, the higher the productivity of the team. Finally, our results suggest that the positive effect of average team familiarity on productivity is amplified when performing more complex tasks.

Our study offers a number of important contributions to operations management literature. Identifying team composition as a major apparatus with which managers can influence productivity; our study provides insights on effective management of fluid team operations. By examining factors including the role of familiarity dispersion, the existence of a bottleneck-pair, whether familiarity is gained from performing more complex tasks, and whether the focal task being performed is a complex one; our study unravels when and how team familiarity influences productivity of fluid teams.

An important focus in the operations management literature has been to increase productivity at the individual, team and organizational level, though for mostly repetitive and routine tasks in contexts such as manufacturing or blue-collar work (Hopp and Spearman 2000, Hopp et al. 2009). A natural extension and focus for this line of work is to study productivity in more knowledge intensive and more complex settings (Hopp et al. 2009). As Peter Drucker famously noted (1999, p. 79) “the most important, and indeed the truly unique, contribution of management in the 20th century was the 50-fold increase in the productivity of the manual worker in manufacturing. The most important contribution management needs to make in the 21st century is similarly to increase the productivity of knowledge work and knowledge workers.” With our focus on fluid team operations in knowledge intensive healthcare settings, and our new insights on team composition strategies and their productivity implications, we believe that our study contributes to operations management literature in a significant way.

In the next section we present the existing literature and develop our hypotheses. We then present our data and results, robustness checks, and finally discuss our findings and conclusions.

3.2 Literature Review and Theory Development

Recent research in operations management has examined operational problems in healthcare settings by addressing issues such as scheduling (Green et al. 2006, He et al. 2012), capacity management (Lee and Zenios 2009, KC and Terwiesch 2012), workload management (Tucker and Edmondson 2003, Powell et al. 2012), and task variety (Avgerinos and Gokpinar 2014). While this paper uses a similar data set to that of Avgerinos and Gokpinar (2015) which studies task variety and learning by focusing on coronary artery bypass grafts, the present study focuses on different research questions concerning team familiarity and related dynamics by analysing the whole set of cardiac surgery operations performed in a hospital. Using the healthcare setting, our work investigates productivity implications of team composition strategies. Because many critical healthcare tasks such as surgeries are performed by fluid teams, individuals’ assignment to these teams is a critical operational decision.

Team familiarity (i.e., prior shared work experience) and its performance implications have been studied in several literatures. First, a number of theoretical studies have suggested positive effects of past collaborations (i.e. team familiarity) on team performance. As team members work together, a transactive memory system (i.e., a set of individual memory systems) in combination with the communication that takes place between individuals (see, Wegner 1986) can be developed among team members (Wegner et al. 1985) which will then improve team productivity. Weick and Roberts (1993) introduced the concept of collective mind-a pattern of affiliations of actions in a social system, and claimed that team stability will increase close relations, which in turn will improve performance. Similarly, shared experience may lead to developing team human capital (Chillemi and Gui 1997) which can also increase team productivity. Also, Edmondson (1999) suggested that team familiarity will improve productivity through promoting team psychological safety.

Second, several studies in healthcare have empirically examined the effect of past collaborations on various performance outcomes. Fleming et al. (2006) conducted interviews with cardiac surgery team members and used survey-based data to show that limited shared experience may lead to breakdowns in communication, decision making and leadership which in turn may threaten surgical safety resulting in adverse patient outcomes. Using data from pediatric, cardiac, and orthopedic operations, Catchpole et al. (2007) also showed that shared experience and effective team working during surgery can prevent minor problems from escalating and becoming serious threats for patients. Finally, Davenport et al. (2007) used survey-data from general/vascular surgery services and found that past collaborations can help decrease patient morbidity.

Broader management literature has generally highlighted beneficial effects of team familiarity on productivity in various empirical settings. Faraj and Sproull (2000) used data from software development teams and showed that familiarity increases productivity since it allows team members to know the area of expertise of each other. Team familiarity also increases the willingness of joint replacements surgery team members to engage in a relationship (Reagans et al. 2005), leading to improved team productivity. Huckman et al. (2009) and Huckman and Staats (2011), show that past common experience can promote team productivity for project teams in the software industry. On the other hand, using data from NBA teams, Berman et al. (2002) argue that easiest gains from shared experience come during the early stage of relationships and after some point this may lead to routinization and lower productivity for teams working in a tacit-knowledge environment. Nonetheless, they point out that practically the number of teams that may experience negative return is extremely low. Katz (1982) also found that stable R&D teams which stay together for more than five years demonstrate lower levels of productivity with compare to teams working together for shorter time. He suggested that high longevity may make team members less motivated, decrease their internal communication and provide them a false assurance.

In addition to examining main effects of average team familiarity, previous studies have also explored the role of several contextual and moderating factors on the relationship between familiarity and performance. Espinosa et al. (2007) used data from software development teams and showed that team size and geographic dispersion interact positively with team familiarity on its beneficial effect on team productivity. Huckman and Staats (2011) focused on the interaction between team familiarity and interpersonal diversity in terms of customer experience, and found that this interaction results in significant improvement in terms of effort and schedule deviation, but it leads to no significant improvement with respect to schedule adherence or project quality. Also, Staats (2012) used data from software development teams to show that horizontal familiarity increase quality whereas hierarchical one promotes productivity.

3.2.1 Familiarity Dispersion

While average team familiarity is an appropriate construct to characterize average levels of prior shared work experience in a team, it may not give us the complete picture with regards to team composition and familiarity dynamics in fluid teams. In fluid team settings in which teams are not stable, there may be high degrees of variation both across and within teams in terms of team members' familiarity among each other. Moreover, two teams with the same levels of average team familiarity may differ substantially in their composition dynamics (e.g., one team with all members moderately familiar with each other, another team in which some pairs highly familiar and other pairs barely familiar with each other). That is, going beyond the examination of the effects of average team familiarity and exploring the role of its dispersion within a team could give us a more nuanced and complete picture in examining the role of familiarity in fluid teams.

In fact, several researchers have pointed out the importance of including dispersion (e.g., diversity) while examining a team-level effect since dispersion has been recognized as a key characteristic for teams in organizations (Harrison and Klein 2007, Jackson et al. 2003). Klein and Kozlowski (2000) suggested that most group constructs come from a combination of mean and dispersion mechanisms. Several researchers have treated mean and dispersion levels as equal properties for the formation of team-level constructs including trust and peer monitoring (De Jong and Dirks 2012), leader-member exchange (Liao et al. 2010) and satisfaction (Dineen et al. 2007). Building on this literature, we identify dispersion in team members' familiarity with each other as an important characteristic of a team, and suggest that "team familiarity dispersion" would be detrimental for team productivity. We argue that team familiarity dispersion may hinder team productivity through two main mechanisms.

First, while past common experience makes individuals aware of each other's expertise and therefore results in more efficient allocation of tasks and knowledge sharing (Edmondson 1999, Faraj and Sproull 2000, Lewis 2003, Reagans et al. 2005), when similar levels of familiarity are not shared by all individuals within the team, they will face difficulties in knowledge sharing and effective integrating due to different beliefs and lack of past collaboration among some members. This will lower productivity (Gardner et al. 2012). Moreover, when relational resources are concentrated within parts of the same team, individuals will develop stronger bonds only with some team members, creating distance from some others (Hornsey and Hogg 2000, Van Knippenberg et al. 2004). This uneven distribution of relational resources may lead to opposition in task-related goals (Bezrukova et al. 2007, Li and Hambrick 2005) or even to conflict and distrust among team members (Van Knippenberg et al. 2007, Choi and Sy 2010). Consequently, teams with high familiarity dispersion will experience decreased productivity because of the unevenly distributed relational resources within the team.

Second, trust among coworkers has been recognized as an important performance driver for healthcare personnel (Cook 2013), and its significance has been highlighted for individuals (e.g. nurses, Altuntas and Baykal 2010) as well as pairs in teams (such as between a surgeon and a nurse, Pullon 2008). Trust is particularly critical for performance in high pressure settings (Colquitt et al. 2011). Cardiac operating room is such an environment (Edmondson 1999, Tucker and Edmondson 2003, Nembhard and Edmondson 2006) where past shared experience will lead to trust development among team members (Gruenfeld et al. 1996, Edmondson 1999). As a result, high familiarity dispersion indicates high trust variance and trust asymmetry within the team, which will have negative effects on team productivity due to unbalanced social exchange structures that inhibit reciprocation of resources. (De Jong and Dirks 2012). Within a team, while individuals are likely to share information and adopt shared information with highly familiar and hence trusted members, they will be less likely to do so with less familiar members (Gruenfeld et al. 2000, Kane et al. 2005, Gardner et al. 2012). This will in turn result in problems in efficient sharing of knowledge within the team. Individuals may refrain from seeking help and advice from less familiar members of the team who may have the right advice and who may be more capable than highly familiar members (Hoffman et al. 2009, Gardner et al. 2012), resulting in reduced productivity. We therefore predict that:

Hypothesis 1*: Team familiarity dispersion has a negative impact on team productivity.*

It is worth noting that our concept of familiarity dispersion is different from interpersonal diversity used in Huckman and Staats (2011). In our case, familiarity dispersion refers to diversity of past common experiences across team members. It focuses on pairs in a team and captures variation in these pairwise familiarity levels in terms of the number of past surgeries performed together. On the other hand, interpersonal diversity, which is insignificant at Huckman and Staats (2011), focuses on individual team members' experience working with specific customers. That is, they capture difference in team members' distribution in working with specific customers.

3.2.2 Familiarity and Bottlenecks

In production settings, bottlenecks are highly important as they constrain the system capacity. They are also fundamental in determining throughput, cycle time, customer service, and other performance metrics (Hopp et al. 2009). Despite its significance in blue-collar settings, and its obvious connotations in knowledge intensive environments, only a small number of studies in the literature have highlighted "bottlenecks" in white-collar settings. Taking the concept of bottleneck to team settings and employing social network analysis, Cross et al. (2001) suggested that some team members may be acting as 'bottlenecks' in sharing information, resulting in the whole team to operate less efficiently. Siemsen et al. (2008) used the motivation-opportunity-ability (MOA) framework to show that one of these three factors can act as a bottleneck and therefore determine the degree of knowledge sharing among members of a workgroup.

Because the smallest unit in a team that exchange information is a pair, and our conceptualization of team familiarity is based on pairwise prior shared work experience, we suggest that the existence of a pair with very low levels of shared work experience (i.e., familiarity) compared to average familiarity of the team may act as a bottleneck for the whole team's operation, and thus reduce productivity. There are several reasons for such an effect. First, emphasizing the role of pairs in team, Bion (1962) suggested that cohesion will develop easier between two individuals than among all group members. One of the initial consequences of prior shared work experience is the development of cohesion (Hackman 1987, Evans and Dion 1991). We argue that the existence of a bottleneck-pair in terms of team familiarity will result in a lack of basic levels of cohesion in that pair, hence constraining some of the members' effective willingness to engage in a relationship and share information within the team, thus negatively influencing team's productivity.

Second, groups and pairs, similar to individuals, can develop habitual routines (Gersick and Hackman 1990) and norms (Feldman 1984). When members have previous common experience, they share a priori premise about the way they should proceed in a given situation (Gersick and Hackman 1990). Bottleneck-pairs which are in significant shortage of habitual routines would need explicit communication about the task itself (Gersick and Hackman 1990, Espinosa et al. 2007) which will in turn reduce productivity of the team.

In addition, normally an important consequence of team familiarity is reducing the possibility of communication errors by enhancing interaction and effective communication among team members, and by reducing uncertainty among them (Harrison et al. 2003). Members who form the bottleneck pair (i.e., with limited familiarity with compare to others) will have uncertainty about each other, which leads to higher anxiety (Gruenfeld et al. 1996) between these members. In a high pressure critical setting, this anxiety may reduce the fluency of individuals (Nemeth 1986), resulting in higher likelihood of communication errors. Because communication errors among surgical team members are the most common cause of operative problems (Makary et al. 2006), and such errors in one pair can be sufficient to cause operative problems and ineffectiveness for the whole surgery, we suggest that the existence of a bottleneck pair will increase the likelihood of communication errors, thus, reduce team productivity.

Finally, cardiac operations consist of a set of steps performed by pairs of the surgical team (KC and Staats 2012) in which tacit knowledge plays a critical role (Edmondson et al. 2003) in smooth performance of these tasks. Because familiarity promotes formation of tacit knowledge among individuals and pairs (Edmondson et al. 2003, Huckman and Pisano 2006) and transferring tacit knowledge is also inherently difficult due to the contextualized nature of the knowledge itself (Joshi et al. 2007), the bottleneck pair will have difficulty in information sharing. Since the entire surgery is a sequence of steps and operations, and the outcome of the whole team depends on the effort of each pair and individual in the team (Hollingshead 2001); the lack of efficient information sharing in one step (i.e, by the bottleneck pair) will have a cascading effect on other steps, which will result in reduced productivity for the whole team. As a result, we expect that:

Hypothesis 2*: The existence of a bottleneck-pair has a negative impact on team productivity.*

3.2.3 Gaining Familiarity

We next consider productivity implications of the way team familiarity is gained. Members in a fluid team inevitably gain familiarity as they perform tasks and work together as part of a team, but in developing this familiarity, the nature of tasks that they perform together could be quite different which may lead to differentiated effects in future task productivity. Consider two similar surgical teams both performing a single operation. Members of both teams will develop familiarity within their teams, but if one team has performed a simple routine operation, whereas the other one has performed a high risk complicated operation, the nature of gained familiarity among team members as a result of performing these operations could be quite different in these teams. The impact of respective gained familiarities in subsequent operations' productivity may be different as well.

As suggested earlier, the development of transactive memory system which involves a representation of who knows what in a team (Lewis 2003) is an essential factor enhancing task performance (Wegner 1986), and we identify it as an important mechanism that links team familiarity and productivity. Brandon and Hollingshead (2004) suggest that experience from more complicated tasks is more beneficial than experience from normal tasks because complicated tasks promote more efficient development of a transactive memory where individuals' resource allocation lead to better encoding, storage, retrieval, and communication of information from different knowledge domains. As a result, we expect team members to develop better transactive memory systems when performing more complicated tasks, and therefore the resulting familiarity to be more beneficial to team productivity in subsequent operations. Individuals working together in more challenging tasks become more aware of expertise of other individuals and are therefore more capable of assigning responsibilities within the team increasing it productivity.

In addition, cognitive interdependence is less likely to occur among team members working on a simple task (Brandon and Hollingshead 2004). In contrast, team members working on a challenging, non-routine task that requires an increased coordination and communication among them, will relate better to each other by forming cognitive interdependence (Levine and Moreland 1999, Moreland 1999). As a result, team members will realize that the outcome of their effort does not depend solely on them but also on the productivity of other group members (Hollingshead 2001). This realization will make each group member more willing and better at sharing information and transferring knowledge among the group (Brandon and Hollingshead (2004), and therefore, becoming more efficient in translating previously gained experience in the past to subsequent tasks, and improve productivity in these tasks.

Finally, cardiac surgeries are characterized by a high level of uncertainty and pressure since the stakes are high (Tucker and Edmondson 2003). Past researchers have also shown that working on simple tasks with low information processing requirements can be decreasingly effective under conditions of uncertainty (Tushman and Nadler 1978, Galbraith 1973). As a result, if gained familiarity comprises of working together on many simple tasks, we expect this to have somewhat limited productivity benefits in performing subsequent tasks. In contrast, if familiarity is gained by performing a higher ratio of complex tasks which inherently require better information sharing and coordination (Gittel 2002), then this will help teams become more productive and cohesive next time they perform a similarly complex or simpler task. Considering all of the above arguments, we predict that:

Hypothesis 3: *The higher the ratio of familiarity which is gained from complex tasks in the past, the higher the productivity of the team.*

3.2.4 Familiarity and Task Nature

Team members who are highly familiar with each other can communicate more effectively since they can refer to the same technical terminology (Cramton 2001), identify and access expertise more effectively (Faraj and Sproull 2000), and generally know what to expect from each individual, and how that can be used to promote team productivity. As a result, when a setting makes it harder for a team to locate specialized knowledge and access expertise, team familiarity can have an increased benefit for the team (Espinosa et al. 2007). That is, we expect the positive relationship between team familiarity and productivity to be amplified when performing more complicated tasks.

Team familiarity provides several significant benefits to team productivity, and these benefits become more prominent during performing more complicated and uncertain tasks. These include reducing the need for explicit interactions (Espinosa et al. 2007), reducing uncertainty about team members (Harrison et al. 2003), and enabling them to figure out more efficient ways to work with each other without the need of extensive interactions (Campbell 1988). Team familiarity also enables team members to deal with variation by applying knowledge gained through their past collaborations (Banker et al. 2001, Field and Sinha 2005), and it may help develop adaptability (Van de Ven et al. 1976, Gittel 2002). We therefore expect team familiarity to have a stronger effect on productivity in environments with high uncertainty that limits communication among team members. Complex operations are characterized by higher levels of uncertainty and they have an increased probability of adverse events (including death). We therefore expect team familiarity to have stronger effects on productivity during these operations. Especially in the operating room, where “you look at the surgeon and you know the body language and you act” (Edmondson et al. 2003, p.221), the influence of team familiarity on productivity will be more prominent during more complicated and therefore uncertain and challenging tasks.

Finally, during a complex operation, it is more likely that complications will arise, increasing the level of complexity of the procedure and making team's task more challenging. During these highly complex and challenging tasks, we expect team members to demonstrate higher cognitive alertness (Levine and Moreland 1999) leading to better motivation and utilization of tacit knowledge, resulting in increased benefits of team familiarity on productivity. As a result, we predict that:

Hypothesis 4: *Team familiarity and task complexity interact positively in their effect on team productivity, such that the positive effect of team familiarity on team productivity is stronger when the nature of the focal task is more complex.*

3.3 Setting, Data and Variables

We test our hypotheses using a dataset of all cardiac surgery operations in a hospital over seven years. Our setting is the cardiac unit of a 300-bed private hospital in Europe, which is property of an American non-profit organization. The hospital serves more than 2,000 patients per year and hosts around 70 cardiac surgeries per month. Our data set consists of archival data of all cardiac surgeries during the period from 01/01/2004 to 31/03/2011 and includes 6,206 operations. For each surgery, the dataset includes, among others, information about the duration of the surgery, specific surgery type, members of the surgical team and patient characteristics.

Each surgery team has three to eight members including the Lead Surgeon, zero to four Assistant Surgeons, the Anesthesiologist, the Perfusionist (a technician who runs the heart-lung machine) and zero to three Scrub Nurses. Our dataset consists of 115 individuals in total: 44 surgeons who appear as Lead or Assistant Surgeon, 12 anesthesiologists, 10 perfusionists, and 49 nurses. 51 of these individuals (19 surgeons, 9 anesthesiologists, 3 perfusionists and 20 nurses) started working after the beginning of our dataset. 37 of the 115 individuals (13 surgeons, 6 anesthesiologists, 4 perfusionists and 14 nurses) do not appear during the last year of our dataset. We also have data regarding in-hospital mortality and information related to the patient's condition prior to operation. Specifically, all patients are divided into three groups by the hospital according to the severity of their case and their characteristics after an initial clinical assessment. These groups are: "Severe", "Medium" and "Mild".

After removing operations for which there are missing data, we are left with 6,171 operations. We also dropped operations during which the patient died since the duration of these operations can be misleading with regards to team productivity. As a result, we are left with 6,129 operations. Using the hospital's classification system, and conducting interviews with several surgeons, we categorize each surgery into one of the following nine major sets: Coronary Artery Bypass Graft (CABG), Valve Repair/Replacement, Congenital Surgery, Heart Failure, Tumor Removal, Routine Cardiac Surgery, Other Normal Surgery (where we include operations that are not characterized by any of the previous categories), Double Surgery (including two types of the previous categories during the same operation) and Triple Surgery (including three previous categories during the same operation). Our dataset consists of 3,273 CABG operations, 1,324 Valve Repair/Replacements, 86 Congenital Surgeries, 70 Heart Failure procedures, 20 Tumor Removals, 185 Routine Cardiac Surgeries, 78 Other Normal Surgeries, 965 Double Surgeries and 128 Triple Surgeries.

Finally, we conduct a limited number of interviews with several staff members to get a better understanding of the cardiac surgery procedures, and to acquire information related to management practice and organizational policies of the hospital.

3.3.1 Variables

Dependent Variable. We use procedure completion time in minutes as our dependent variable. Consistent with prior research (Pisano et al. 2001, Edmondson et al. 2003, Reagans et al. 2005) we contend that lower completion times represent an increase in the productivity of the surgical team. The duration of the procedure is considered to be one of the most important factors affecting the total costs for treating patients undergoing a cardiac surgery (Pisano et al. 2001). In our dataset, the average duration of an operation is equal to 294 minutes, with a minimum of 29 minutes and a maximum of 900 minutes.

One potential issue with regards to our dependent variable is that lower completion times might be argued to have a negative impact on the clinical outcome of the surgery. However, after interviewing hospital staff and analysing our data, we have concluded that this is not the case. Specifically, there is no correlation between in-hospital mortality and the duration of the operation. Also, other than the clinical condition of the patient before the operation (i.e, severe), none of the variables have significant association with in-hospital mortality. In addition, we also investigate for any systematic statistical relationship between in-hospital mortality and our dependent (i.e. duration) and all independent and control variables (e.g., those at patient, surgeon, or team level). In order to examine this, we group our severe operations into two, those where in-hospital mortality occurred, and those where it did not. We then conduct a t-test to examine any potential relationship between in-hospital mortality and our independent and control variables. Our results indicate that there is no statistical difference between these two groups (i.e., mortality vs no-mortality group) in terms of the averages of the variables of interest such as duration, familiarity, etc. Finally, we have examined for any association between surgery durations and surgeon assignments. In order to check this, we grouped our surgeries into four categories according to their durations by using 25th percentile, the median, and 75th percentile as cut-offs, and then examined lead surgeons' assignments to these groups with a chi-square test. Our results indicate that they are evenly spread.

Moreover, prior research has shown that lower completion times can actually improve the quality of the operations outcome (Pisano et al. 2001) whereas longer surgical durations have been associated with higher probabilities of a surgical site infection (SSI) in several surgery types including cardiac operations (Gaynes et al. 2001, Gibbons et al. 2011). As a result, in contrast to the argument suggesting that lower completion times may be associated with worse clinical outcomes; previous research suggests that they are actually associated with better clinical outcomes.

3.3.1.1 Independent Variables

Team Familiarity Dispersion. In order to capture team familiarity dispersion, we first calculate the number of past collaborations for all pairs within the team. This represents pairwise familiarity for each pair in a team. We then calculate the standard deviation of these pairwise familiarity scores. This way, we characterize familiarity dispersion within the team, and we distinguish between teams which may have similar levels of average team familiarity, but with a difference in the distribution of past collaborations among the pairs in a team.

Notice that dispersion can be operationalized in several ways depending on the theoretical construct. Indeed, Harrison and Klein (2007) provide a related discussion of three major types of diversity constructs (i.e., "separation", "variety" and "disparity) and their appropriate operationalization. Our conceptualization of familiarity dispersion within fluid teams where pairs differ from one another in their position along a continuous attribute (e.g., team familiarity) is very similar to their "separation" construct. Because we are not interested in differences among team members on a categorical attribute, "variety" is not an appropriate construct for us. Similarly, as we are not interested in the portion possessed or share of a desired resource, "disparity" is not a suitable construct for our theoretical perspective on the dispersion of team familiarity. In short, in our conceptualization and operationalization of team familiarity dispersion, we concur with Harrison and Klein (2007)'s distinction both in terms of construct type and measurement, and use standard deviation which is the most common way of capturing "separation" type constructs like ours. Notice that we have also considered alternative measures (e.g., coefficient of variation, range) for team familiarity dispersion in the robustness checks section.

Bottleneck-Pair. For every operation, we first identify the pair with the minimum number of past collaborations in the team (i.e., the least familiar pair), and get its pairwise familiarity score. We take this number and divide it by the average familiarity of that team. We call the resulting number "bottleneck score" for the team, which is a proxy for how low the familiarity is in the least familiar pair with compare to team's average familiarity. We then use a set of discrete "buckets" of bottleneck scores for our main model. Specifically, we create a new variable which is equal to 4 if the bottleneck score is lower or equal to 25th percentile, 3 if it is between 25th percentile and the median, 2 if it is between the median and 75th percentile and 1 if it is above or equal to 75th percentile. With this approach, we assign each bottleneck score to a bucket which helps us identify those teams who are more susceptible to bottleneck effects (i.e., those with very low bottleneck scores -lower than 25th percentile-) versus those who are less susceptible to such bottleneck effects (i.e., those with bottleneck scores in other buckets).

Another potential way of capturing this effect could have been to take the pair with the lowest familiarity score in every team and use their pairwise familiarity scores directly, without dividing by the average team familiarity. Notice that similar to a factory setting, we expect pairs with the lowest familiarity scores to act as bottlenecks relative to the rest of the pairs within the same team. That is, we argue that existing levels of information sharing and collaboration within a team will be hampered by a pair whose familiarity is much lower than the team's average. Hence, taking the pair with the lowest familiarity in every team and using their familiarity score directly would not have been appropriate as a low score may merely be an indication of a newly formed team with many unfamiliar pairs rather than a bottleneck.

One could have also used bottleneck scores (lowest score divided by team average) directly as a continuous variable. Notice that, instead of doing so, we assign each team to one of the four buckets described above depending on their bottleneck scores and use these in our model. This is for two reasons. First, bottleneck score as a continuous variable is extremely highly skewed (see Figure 3.2 of Appendix). As a result, even transforming the variable (e.g., log transformation) would not alleviate this extreme skewness problem. Second, our conceptual model suggests that the existence of a bottleneck pair in a team would hinder overall knowledge transfer capability of a team. However, we do not expect this effect to be continuous and linear. Instead, our model suggests that certain teams are more susceptible to bottleneck effects than others. For example, consider two teams with very high bottleneck scores (i.e., their least familiar pairs still do have a good level of familiarity compared to team's average). Then, even if these two teams differ in their actual bottleneck scores, we expect both of them to demonstrate no bottleneck related hindrance, that is, we do not expect to see any difference in their productivity as a result of difference in their bottleneck scores. That is, using a continuous variable for bottleneck scores would have been misleading.

Finally, notice that we have considered alternative measures of this variable in the robustness checks section.

Severe Familiarity Percentage. In our setting, each procedure is characterized as “severe”, “medium” or “mild” according to the condition of the patient by the hospital after an initial clinical assessment. We claim that “severe” cases constitute more complex tasks. Schroder et al. (1967) identified information load, information diversity and degree of uncertainty as three properties of a complex task and showed that complexity increases as each of these properties increases. In a highly uncertain task the desired outcome is not clearly linked to specific actions, making therefore the task more complex and challenging (Campbell 1988). In addition, in a healthcare setting like ours, input uncertainty, which refers to the uncertainty to the input of a production process, exists due to different patient conditions (Argote 1982, Gittel 2002). Severe cases generally constitute more challenging tasks since they are characterized by an increased probability of adverse events (including death) and therefore higher uncertainty. Hence we break down team familiarity into two components: Familiarity gained from cases in which the patient’s condition is characterized as “severe” and familiarity gained from cases in which the condition is characterized as either “medium” or “mild” and characterize the severe ones as complex tasks.

This variable captures the ratio of team familiarity that is gained through working together in severe (i.e. more complex) cases. In order to calculate this variable, we first calculate the number of times every pair in the team has worked together before (excluding the current operation) and take the sum for all pairs in the team. Next, we count the number of times every pair in the team has worked together before in cases in which the condition of the patient is characterized as “severe” and take the sum for all pairs in the team. Then, to get Severe Familiarity Percentage, we divide the latter number with the former one. This way, we capture the percentage of past collaborations that took place during a complex operation and expect it to have a beneficial effect on team productivity.

3.3.1.2 Control Variables

Average Team Familiarity. Similar to previous works (Reagans et al. 2005), we count the number of times every pair in the team has worked together before (without including the current operation), take the sum for all pairs in the team and then divide this number by (N(N-1)/2), where N is the team size. This variable represents the average level of familiarity within the team. Similar to previous empirical studies (Reagans et al. 2005, Huckman et al. 2009, Huckman and Staats 2011, Staats 2012) we expect team familiarity to promote team productivity.

Team Size. We control for the number of all team members N. Larger teams may have difficulty in coordination, which could result in decreased productivity (Hackman 2002, Reagans et al. 2005).

Individual Average Direct Experience. We control for individuals' prior experience in the same type of procedures. For each team member, we calculate the number of times they appear in a same type procedure prior to the current procedure (not including the current one). We then take the sum, and divide by the number of team members. This way, we control for individual experience of team members in a same type of operation. Instead of considering individuals' direct experience only in the same type of operations in the past, we could have also considered their experience on other types of operations (i.e., indirect experience), and combine them in a variable called Individual Average Experience which captures the total number of operations participated by each team member prior to the focal one. Our results remained the same when our present variable is replaced with this one. In addition, we control separately for Lead Surgeon's experience and the other team members (since one can claim that Lead Surgeon's experience is more important than the one of the other team members) and the results remain the same.

Indicator of Quarter. In order to capture any effects of potential technological advances that may influence operation times, or potential changes in the hospital policy that may have an impact on operation durations (e.g., additional paperwork required, etc.), we include an indicator variable indicating the quarter that the operation was conducted.

Indicator for Patient Condition. In order to control for the patient's condition which is characterized as "mild", "medium", or "severe", we include two dummy variables in all models. Considering "medium" as the reference category, "severe" variable is equal to one if the case is characterized as severe and zero otherwise, whereas "mild" variable is equal to one if the case is characterized as mild and zero otherwise. Because doctors already take into account a good number of critical factors such as age, clinical history, past complications, other existing health problems, etc. when they classify patients into these three categories prior to the operation, we believe it is a good indicator of overall patient characteristic.

Indicator for Procedure Type. As described above, each operation can be characterized and considered in one of the nine categories. Hence, in our model we include eight indicators (we consider the most common type CABG as the reference category) that are equal to one if the procedure is in the respective category and zero otherwise.

Indicator for the Lead Surgeon. In order to control for the skill of the Lead Surgeon, which may be one of the most influential factors in surgery performance, we control for the Lead Surgeon fixed effects by including a dummy variable indicating the Lead Surgeon in each operation.

Indicator for the Day of the Week. Similar to previous research in cardiac surgery settings (KC et al. 2012) we include a dummy variable indicating the day of the week that each operation was performed.

3.4 Results

Given the structure of our data and the level of our analysis, to test our hypotheses, we use Ordinary Least Square Regression with AR(1) covariance structure to control for serial correlation among operations performed close in time (see Reagans et al. 2005 for similar structure). We also checked for normality of residuals and for heteroscedasticity. Also, consistent with the previous research which introduced team familiarity using a learning curve model (Reagans et al. 2005) and the learning literature (Pisano et al. 2001, Edmondson et al. 2003), we similarly take the logarithm of our dependent variable (i.e., procedure completion time) and our continuous variables which are learning related and cumulative in nature over time such as average team familiarity and individual average direct experience. However, for other variables which are not cumulative in nature such as team familiarity dispersion and team size, we keep their original forms. It is also worth noting that we checked the distribution of all our variables with ladder and gladder commands in Stata, and log transform our variables accordingly. Our model is the following:

$$\begin{aligned} ln(\textit{procedure completion time}_t) = \; & \beta_0 + \beta_1 (\textit{team familiarity dispersion}_t) + \\ & \beta_2 (\textit{bottleneck-pair}_t) + \\ & \beta_3 \ln(\textit{severe familiarity percentage}_t) + \\ & \beta_4 (\textit{severe}_t) \times \ln(\textit{average familiarity}_t) + \\ & \beta_5 \textit{ controls} + u_t , \\ & \textit{where } \; u_t = \rho u_{t-1} + e_t \end{aligned}$$

Table 3.1 shows descriptive statistics and correlations among the variables. It reports the values for the logged and normalized variables as described earlier. Table 3.2 shows the results for hypotheses 1, 2, 3 and 4. Notice that we report robust standard errors in all our models. Also, while our final sample includes 6,129 operations, notice that the number of observations in Table 3.2 is equal to 6,128 due to the AR(1) covariance structure. In order to check for any potential multicollinearity issues, we introduce our variables of interest one at a time in our main model. In addition, we checked the Condition Index (CI) and it is below the suggested threshold of 30 (Cohen 2003). In model 1, we include only the control variables. In model 2, we include *Team Familiarity Dispersion* and we observe a 0.84% increase (significant at 1% level) in the adjusted R^2. AIC and BIC also indicate that model 2 is superior to model 1. Team Familiarity Dispersion has a positive and significant coefficient at 1% level, providing support for our first hypothesis. In model 3 we remove *Team Familiarity Dispersion* and include *Bottleneck-Pair.* We observe a 0.84% increase (significant at 1% level) in the adjusted R^2 compared to model 1 and both AIC and BIC also indicate that model 3 is superior to model 1. The coefficient of the Bottleneck-Pair is significant at 1% level and positive providing support for H2. In model 4 we include both *Team Familiarity Dispersion* and *Bottleneck-Pair* and still find support for both H1 and H2. We now observe a 1.49% increase (significant at 1% level) in the adjusted R^2 compared to model 1 and both AIC and BIC indicate that model 4 is better than all previous models. In model 5 we add *Severe Familiarity Percentage* and we observe a 0.41% increase (significant at 1% level) in the adjusted R^2 compared to model 4 and both AIC and BIC indicate that model 5 is superior to model 4. *Severe Familiarity Percentage* is significant at 1% level with a negative coefficient. That is, Hypothesis 3 is also supported.

While the increase in the variation explained (adjusted R^2) between the initial model (with only control variables) and the full model may be small (i.e., 1.89%), this is not surprising. This is because of the dominance of clinical factors in explaining variation in most healthcare performance metrics. Indeed, similar studies investigating organizational or operational phenomena in healthcare settings (e.g., KC and Terwiesch (2011), KC and Terwiesch (2009), Reagans et al. (2005)) report similarly small changes in adjusted R^2 between their initial models and full models.

An increase of one standard deviation on Team Familiarity Dispersion will increase the duration of the operation by 2.14% (assuming all other variables remain constant). In addition, the coefficient of our variable Bottleneck-Pair is equal to 0.0176, which suggests that moving from one category to another increases the duration of an operation by 1.76% (assuming all other variables remain constant). Furthermore, an

3.1 Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9
1. Duration	5.65690	0.268251	3.367296	6.768493	1								
2. Team Familiarity Dispersion	278.1966	213.7607	0	1097	0.0902**	1							
3. Bottleneck-Pair	2.50057	1.118052	1	4	0.0509**	-0.0267*	1						
4. Severe Familiarity Percentage	0.090134	0.031086	0	0.430783	-0.0649**	-0.3273**	0.2116**	1					
5. Average Team Familiarity	4.990338	1.030308	0	6.928375	0.0222	0.6274**	0.2910**	-0.4432**	1				
6. Individual Average Direct Experience	5.075506	1.376165	0	7.226936	0.2419**	0.5939**	0.0906**	0.2458**	-0.2686**	1			
7. Team Size	5.414456	0.785237	3	8	0.0442**	0.0519**	0.3323**	0.0437**	0.0104*	-0.0734**	1		
8. Severe	0.183061	0.22417	0	1	-0.0925**	-0.0365	0.0536*	-0.0907**	0.0638**	-0.2034**	0.0670**	1	
9. Mild	0.220762	0.420715	0	1	-0.0185	-0.0136	0.0157	-0.0165	-0.0162	0.0419**	-0.1159**	-0.1793**	1

+, * and ** denote significance at 10%, 5% and 1% levels respectively.
Logged values for all variables except Severe and Mild

Table 3.2 Regression of Team Familiarity on Surgery Duration

Variable	Duration					
	Model: (1)	(2)	(3)	(4)	(5)	(6)
Team Familiarity Dispersion		0.0001**		0.0001**	0.0001**	0.0001**
		(0.0000)		(0.0000)	(0.0000)	(0.0000)
Bottleneck-Pair			0.0113**	0.0170**	0.0176**	0.0178**
			(0.0032)	(0.0036)	(0.0036)	(0.0036)
Severe Familiarity Percentage					-0.4106**	-0.4087**
					(0.1141)	(0.1152)
Average Team Familiarity	-0.0290**	-0.0354**	-0.0390**	-0.0413**	-0.0460**	-0.0425**
	(0.0092)	(0.0120)	(0.0099)	(0.0141)	(0.0142)	(0.0142)
Average Team Familiarity x Severe						-0.0284*
						(0.0124)
Individual Average Direct Experience	-0.0676**	-0.0675**	-0.0703**	-0.0515**	-0.0470**	-0.0481**
	(0.0114)	(0.0134)	(0.0124)	(0.0144)	(0.0146)	(0.0145)
Team Size	0.0386**	0.0379**	0.0418**	0.0458**	0.0462**	0.0469**
	(0.0060)	(0.0061)	(0.0064)	(0.0065)	(0.0065)	(0.0066)
Severe	0.0243**	0.0237**	0.0250**	0.0243**	0.0243**	0.1614**
	(0.0043)	(0.0043)	(0.0043)	(0.0043)	(0.0043)	(0.0600)
Mild	-0.0124*	-0.0129*	-0.0112+	-0.0119*	-0.0118*	-0.0114+
	(0.0060)	(0.0060)	(0.0060)	(0.0060)	(0.0060)	(0.0060)
Constant	5.3009**	5.3158**	5.2591**	5.2637**	5.2985**	5.2812**
	(0.0812)	(0.0806)	(0.0822)	(0.0819)	(0.0825)	(0.0827)
Observations (N)	6128	6128	6128	6128	6128	6128
Adjusted R^2	0.477	0.481	0.481	0.484	0.486	0.487
Increase in Adjusted R^2		0.004**	0.004**	0.003**	0.002**	0.001*
BIC	-1512.742	-1518.935	-1517.899	-1531.114	-1533.067	-1535.958
AIC	-1700.992	-1707.909	-1712.873	-1732.811	-1741.482	-1751.096
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Procedure Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

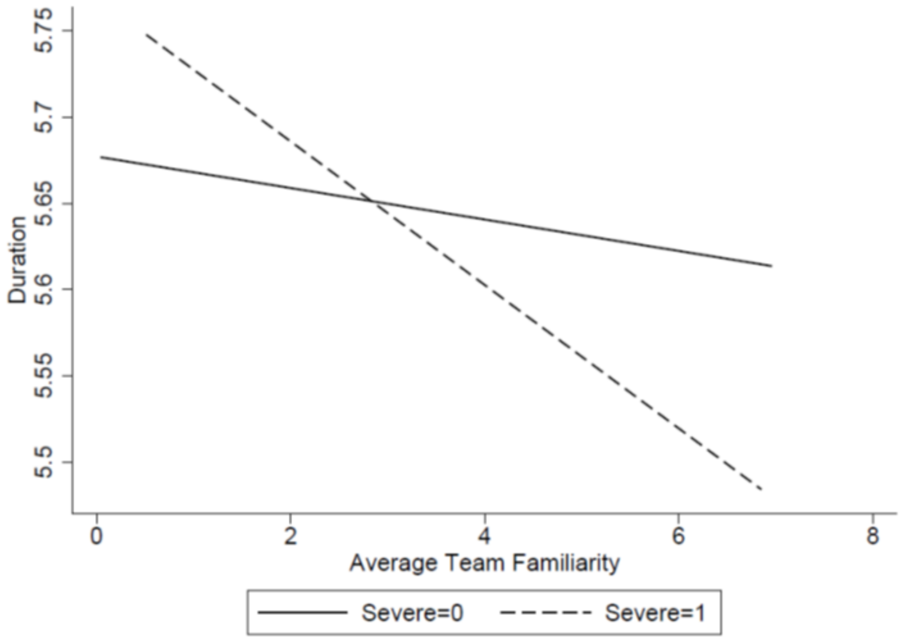
+, * and ** denote significance at 10%, 5% and 1% levels respectively

increase of one standard deviation on Severe Familiarity Percentage will decrease duration by 16.84% (assuming all other variables remain constant).

While some of these percentages may appear small, considering the fact that in the hospital hundreds of operations take place in a year, the aggregate effect would be quite substantial. In addition, as we have mentioned earlier, lower completion times are shown to be associated with better clinical outcomes. We believe therefore that even a slight decrease in the duration of an operation can improve the surgical outcome and reduce the probability of an adverse event during or after the surgery.

In order to test our fourth hypothesis, we include an interaction term, which is equal to the product of the dummy variable *Severe* and the continuous control variable *Average Team Familiarity*. Model 6 of Table 3.2 presents the results for our fourth hypothesis. We observe that the results for our previous hypotheses remain the same and the adjusted R^2 is increased by 0.21% (significant at 5% level). The interaction term is significant at 1% level and negative, providing support for H4. AIC and BIC also indicate that model 6 is superior to model 5. Finally, we conduct a post hoc plot for Hypothesis 4 in order to examine the moderating effects as described by Aiken and West (1991) and Dawson and Richter (2006). Figure 3.1 reveals the moderating relationship as suggested in Hypothesis 4.

Figure 3.1 Interaction Plot for H4



3.5 Robustness Checks

We perform a number of robustness checks to check the sensitivity of our results and to rule out potential alternative explanations. We group these robustness checks into four sections as listed below:

3.5.1 Potential Biases related to Data

While we observe team familiarity over a long time period (i.e. seven years) and in large number of operations, we do not have information about team assignments prior to our dataset, that is, there is some unobserved familiarity before the start of our dataset. Similar to Reagans et al. (2005) which use a very similar data structure to calculate team experience, we expect our results to have no systematic bias. Nevertheless, we investigate potential effects of missing team familiarity by repeating our analysis after excluding different time intervals from the beginning of our dataset. That is, we re-run our analyses in subsamples by removing the first 12 months and 24 months respectively from our overall dataset of 87 months. In each subsample, the observations include only surgeries in that subsample and we also calculate team familiarity by using that subsample (e.g., in the removal of 24 months case, our observations include only those surgeries between month 24 and month 87 and when calculating team familiarity we also use surgeries between month 24 and month 87). This way, we investigate the robustness of our results when there is missing data at the beginning. Our results are the same qualitatively (with different magnitudes as expected) both in the full sample and in the subsample (where there is simulated missing experience), providing further support for our hypotheses.

We also check the sensitivity of our results by repeating our analysis differently. In this case we remove the first 12 and 24 months, and while the observations include only surgeries in the remaining subsample, we calculate team familiarity by using the whole dataset. (e.g., when removing the first 24 months, our observations include only those surgeries between month 24 and 87, but when calculating team familiarity we use all previous surgeries from month 0). Again our results are the same qualitatively.

3.5.2 Potential Biases related to Methods and Variables Choice

While our log transformation of the dependent variable is consistent with the literature, this is equivalent to a duration or survival model with a lognormal distributional assumption (Cameron and Trivedi 2005, Cleves et al. 2010). This may put a strong functional form assumption on the shape of the hazard function which may influence our results. Therefore, in order to check that our results are not sensitive to linear or log specification of our dependent variable, we have re-run all our analyses using linear form of the dependent variable. All of our results remained the same in this specification. We also examine any non-linear effects of average team familiarity and familiarity dispersion on team productivity by adding their squared terms into our models and found that they were insignificant as expected.

A potential issue with our use of AR(1) covariance structure is that our sample does not satisfy the equal-spacing assumption. We believe that this does not introduce any significant problems to our results because we have a large sample (i.e., 6,206 observations) representing a long time interval (i.e., 87 months). Nonetheless, we repeated our analysis after controlling for the unequal spacing in our dataset. Specifically, for each operation we calculate the time difference between subsequent operations in hours, and include this new variable in our analysis. Our results remained the same for all of our hypotheses. In addition, although our AR(1) approach is consistent with the previous literature and Durbin-Watson test suggests first-order autocorrelation, one may argue that sequential relationship may not hold for certain operations in our case (if they somehow overlap), or serial correlation may be of higher order. We performed two sets of additional analyses for these. First, we repeat our analysis without an AR(1) covariance structure, and our results remained the same. Second, we assume serial correlation may be of higher order and repeat our analyses with AR(2), AR(3), and AR(4) structures. In all models, our results remained unchanged.

We also consider alternative measures of our main independent variables. First, instead of using *standard deviation* to capture team familiarity dispersion, we consider *coefficient of variation* (i.e., standard deviation/mean) and *range* (i.e., max-min) of pairwise familiarity scores in our teams. Our results with these alternative measures remained qualitatively the same. Second, we consider an alternative construction of our bottleneck-pair variable. Specifically, after calculating the bottleneck score as before (i.e., familiarity of the least familiar pair/average familiarity of the team), we create four categories/buckets depending on the bottleneck score: (i) if the bottleneck score is lower than or equal to 25th percentile, (ii) if it is between 25th percentile and the median, (iii) if it is between the median and 75th percentile, and (iv) if it is higher than or equal to 75th percentile. We then used three indicator (dummy) variables for these categories taking the last category as reference and re-run our models. Indicator for the first category (i.e., i) is significant at 1% with a positive coefficient, indicator for the second category (i.e., ii) is significant at 5% and positive with a lower coefficient than the first category (a t-test also indicated that the two coefficients are different), and the indicator for the third category (i.e., iii) is insignificant. This confirms our earlier results for the second hypothesis. In addition, we use an alternative continuous variable for Bottleneck-Pair, which is equal to the absolute distance between the least familiar pair and team's average familiarity. We find that this variable is significant and positive, again supporting our H2. Finally, we also use alternative variables for both dispersion and bottleneck-pair in the same model (*coefficient of variation* and absolute distance between the least familiar pair and team's average familiarity, and *range* and absolute distance between the least familiar pair and team's average familiarity) and get the same results for H1 and H2.

3.5.3 Potential Selection Biases

Selection of team members in surgery operations could be a significant concern. Although according to the hospital management the assignment of surgeons is random, Lead Surgeons may prefer working with specific Assistants Surgeons, and one may argue that this may influence our results related to Hypothesis 1. In order to address this, we repeated our analysis by calculating team familiarity dispersion without taking into account any pairs that include only surgeons (i.e., surgeon-surgeon pair) in a team. Our results with modified team familiarity dispersion remained the same qualitatively (with a reduction in the magnitude as expected). In addition, nurses are randomly assigned to surgery operations according to their shifts, and anaesthesiologists and perfusionists are randomly assigned by the hospital for each operation. Nonetheless, in order to eliminate any potential bias in our results we examined this issue further: We calculated for each Lead Surgeon the percentages of the operations that she has worked with each anaesthesiologist, perfusionist and first scrub nurse (i.e. lead nurse). In anaesthesiologists, the highest percentage is 33.10% (but the Lead Surgeon has worked with 9 out of the 12 anaesthesiologists in our dataset), in perfusionists the highest percentage is equal to 36.19% (and the Lead Surgeon has worked with 6 out of the 10 perfusionists appearing in our dataset) and in lead nurses it is equal to 29.81% (the lead surgeon has worked with 38 out of the 41 nurses appearing as lead nurses in our sample). Considering these percentages and the fact that out of the 115 individuals in our sample, 51 of them (19 surgeons, 9 anesthesiologists, 3 perfusionists and 20 nurses) started working after the beginning of our dataset and 37 (13 surgeons, 6 anesthesiologists, 4 perfusionists and 14 nurses) do not appear during the last year of our dataset, we believe that there is no systematic selection of lead nurses, anesthesiologists, or perfusionists by the Lead Surgeon in the hospital.

One may also argue that some surgeons could possibly choose operations to avoid patients with high chance of complications (e.g., death), which may bias our results. Similarly, more severe cases might be assigned to more experienced Lead Surgeons, or easier cases might be assigned to junior Lead Surgeons. First, according to our interviews, official hospital policy is clear in not allowing surgeons to choose or avoid certain operations. Second, we investigate the spread of “severe”, “moderate” and “mild” operations among surgeons with a chi-square test of independence where the null hypothesis is that each patient’s assigned doctor is independent of her condition. We reject the null hypothesis (with chi-square statistics of 13.271, 16.887 and 14.674 for each operation type respectively), and conclude that surgeries are indeed evenly spread among lead surgeons. Third, in our analysis we have excluded all highly complex cases that resulted in the death of the patient during surgery (42 in total). We repeat our analysis after including them and get the same results in terms of significance. Finally, we repeat our analysis after excluding the severe and all the mild cases in our sample, and left with only the medium cases. Our results for Hypotheses 1, 2 and 3 remained the same qualitatively.

Finally, if lead surgeons prefer to work with more familiar people for severe cases (since there is an increased likelihood of major adverse events like patient death), this may create bias in our results for H4. In order to rule out this alternative explanation, we examined an alternative specification of task complexity. We created a new variable called *Patient*, which reflects the condition of the patient before the operation. Recall that severe cases represent the most challenging and complicated ones. In the new specification, *Patient* is equal to 1 if the condition of the patient characterized by the hospital is “mild”; 2 if it is “medium”; and 3 if it is “severe”. Next, we created a new interaction term by multiplying *Average Team Familiarity* and *Patient* and we ran our analysis for H4 using the new variables. Our results for hypotheses 1, 2 and 3 are the same in terms of significance and very similar in terms of coefficients, and the new interaction term is also significant and negative at 5% level providing support for the fourth hypothesis. We therefore believe that our results for H4 are not biased by a potential selection of more familiar members from the lead surgeon in severe cases. This is because we observe an increasing beneficial effect of team familiarity on productivity as the complexity of the task being performed is increased (i.e., from mild-1 to severe-3).

3.5.4 Other Alternative Explanations

An alternative explanation for H1 and H2 could be that the detrimental effects from familiarity dispersion and bottleneck-pair may actually be created not by the mechanisms we suggested, but rather by the sole presence of a new team member or a new pair (i.e., new to his/her teammates), or by a team member with very little prior individual experience (i.e., new to the hospital). In order to address this, we conducted two analyses. First, we ran our analysis after removing 780 operations from our sample in which there is a new member, or a pair working together for the first time. Second, we repeated our analysis after introducing a new variable which is equal to the direct experience of the team member with the lowest direct experience in the team. Our results remained qualitatively the same in both cases.

An alternative explanation for H3 could be that it is not familiarity, but individual experience gained in highly complex operations in the past that may make the team more productive subsequently. In order to test this, we repeated our analysis for H3 after replacing the variable Individual Average Direct Experience with another variable which captures the individual experience of team members from severe operations in the past, and we called this Individual Average Severe Experience. It is calculated the same way as Individual Average Direct Experience, this time by only counting severe cases though. Introducing this new variable did not change our conclusions for H3. Moreover, we repeat our analysis after controlling for the Lead Surgeon Direct Experience (while including the same variables for other team members too) and find the same results.

3.6 Discussion and Conclusions

Our study extends literature in fluid team operations by examining the role of team composition dynamics (i.e., team familiarity) on team productivity through detailed and nuanced mechanisms. Although organizations increasingly rely on fluid teams to perform essential tasks, our understanding of formation and operations of fluid teams remain limited. In this study, we make use of a large scale dataset that involves all cardiac operations performed in a hospital over more than seven years, which allows us to observe the relationship between team composition dynamics and its productivity implications during a long time period. Studying fluid teams at a granular level, we offer a micro-founded account of how shared work experience among team members influence team productivity.

Most of the previous literature in operations and management has focused on mean levels of familiarity within teams (e.g., Edmondson 1999, Faraj and Sproull 2000, Reagans et al. 2005, Espinosa et al. 2007, Huckman et al. 2009, Huckman and Staats 2011) and has largely ignored the distribution of pairwise familiarity scores. This is a significant limitation for understating the role of team familiarity in fluid settings, because fluid teams have a dynamic nature in terms of team memberships, and there is likely significant variation in terms of pairwise familiarity within these teams even if they have similar average levels of team familiarity. Consequently, the present study improves our conceptual understanding of team familiarity by considering not only the mean, but also the shape of the familiarity distribution, and by demonstrating the importance of dispersion and bottlenecks in explaining team productivity. In addition, considering the two levels of relationships in fluid team settings: At the pair and at the team level; while the emergence of average team familiarity is a straightforward extension of pairwise familiarity, this is not the case for dispersion and bottlenecks. That is, team familiarity dispersion and bottleneck enriches our understanding of familiarity because they capture distinct team level characteristics that are undefined at the individual or pair level and they only emerge at the team level. In addition, although these two are somewhat related, we contend that dispersion and bottleneck are two distinct theoretical concepts. While the existence of a bottleneck pair (i.e., a pair with very low familiarity compared to the average familiarity of the team) may indicate high dispersion, the opposite is not true. A high dispersion can be caused by one or two pairs that are highly familiar compared to the other pairs in the team. This does not necessarily mean that there will be a bottleneck pair. They are therefore distinct concepts that capture two important characteristic of familiarity distribution within a team.

We believe the new refinements of team familiarity proposed in this paper are not only statistically significant but also managerially significant, as our results suggest new and potentially different managerial actions than those previously suggested in the literature. For example, finding a positive effect of average team familiarity on performance, Huckman et al. (2009) and Huckman and Staats (2011) suggest “the managerial advice to keep teams together” (p.325) to increase the average familiarity of the teams. However, they also acknowledge the practical difficulty of such strategy as "all team members cannot work on teams with high team familiarity" (p.325). In our study, our key findings highlight negative effects of team familiarity dispersion and bottlenecks pairs on productivity. Therefore, rather than trying to keep team members together by allocating the same team members to the same teams repeatedly (which is also not practical due to factors such as employee turnover and newcomers, a common occurrence in many fluid team settings), we suggest managers to actively focus on the least familiar pairs and individuals and try to increase their familiarity levels, and discourage pairs who already have very high levels of familiarity with compare to others. In order to assess practical implications of our study, we next run a basic policy simulation with simple team assignment rules based on our findings. For this simulation, we used real data from our study by picking a month after the middle of our data interval so that we can observe diverse set of familiarity scores (e.g., month 49, the first month at the beginning of year 5, which also has no newcomers for simplicity purposes). For the policy, we use existing pairwise familiarity levels (i.e., those from the beginning of our dataset to the focal month), and form a new team for each upcoming surgery during the next month based on the proposed rule. We then calculate any potential time savings from the proposed team assignment rule based on our econometric model. Specifically, for each available individual (excluding the Lead Surgeon) we calculate their overall familiarity score from all surgeries with all pairs in the past. Then for a team of N members (excluding the Lead Surgeons, who are assigned on a rotating schedule) we assign the N individuals with the lowest score while prohibiting to include the most familiar pairs of our sample. This way we were able to decrease dispersion for that specific month by around 3 standard deviations. Team familiarity dispersion was reduced by 32.34% (277.55 units, which indicates that duration will be decreased by 2.78%). Bottleneck score was increased by 93.73% (which indicates that duration will be decreased by 5.28% since our categorical variable will move from the highest category, equal to 4, to the lowest one, which is equal to 1). Average team familiarity was also decreased by 27.02% which indicates that duration will be increased by 1.24%. Taken together, this new team assignment policy decreases surgery durations by 6.82% (20.05 minutes). This reduction corresponds to conducting 58 additional operations per year.

An important observation we had is that a pair within the team with a familiarity level much less than the average familiarity of the team may slow down the whole team, function as a bottleneck, and cause delays and eventually decrease the productivity of the whole team. Our conceptualization of bottleneck-pair could also be viewed with the lens of recent research about the introduction of surgical checklists and their benefits

(Gawande 2010). A central feature of surgical checklists is preoperative team briefings where each person in the team speaks and introduces herself to other team members. Surgical checklists are suggested to be highly significant in improving teamwork and communication in the operating room (Gawande 2010), and team briefings have been associated with better surgical outcomes (Lingard et al. 2008). We suggest that, a key benefit of team briefings could be to reduce or eliminate the potentially negative effect of a bottleneck pair in team. While team briefings may not contribute much to team members who are already quite familiar with each other, it may provide substantial benefits in reducing potential communication and information sharing problems between team members with very little prior shared work experience (i.e., bottleneck pairs). Our results indicate that the existence of such a bottleneck pair can increase the duration of an average operation by 5.17 minutes. So eliminating such a bottleneck pair (moving a pair from the highest category to the next one) will result in about 6-hours savings in total surgery time per month. Also, when there is a new staff member, hence the effect is inevitable; managers can moderate the detrimental effect by initially assigning her and other relatively new employees to simple cases together. Future research could also investigate the effect of a bottleneck pair based on different individuals with different roles within the same team. In our setting for instance a bottleneck pair consisting of two surgeons might have a more detrimental effect than one consisting of a surgeon and a nurse. With our current approach there is no distinction between these two bottleneck pairs. Such a distinction could also shed more light on how bottleneck behaviours interact with the different roles of unfamiliar individuals within a fluid team.

Another important finding which may help managers to develop better team composition strategies is that team members that have worked together previously in more challenging tasks could develop a transactive memory system more effectively. Similarly, they may create stronger bonds and relationships when their shared experience comes from complicated cases, which helps them better collaborate in subsequent tasks, therefore increasing their future productivity. That is, when assigning individuals to teams, managers could consider not only the amount of shared work experience, but also the nature of shared experience in past operations (e.g., in challenging vs. less challenging tasks). Consider two teams with the same amount of shared work experience, where members of the first team have performed equal number of severe and non-severe operations, and hence Severe Familiarity Percentage is 0.5, and members of the second team who performed only 10% higher amount of severe operations with compare to the first one, and hence Severe Familiarity Percentage is 0.55. The second team will need 4.11% less time (12.08

minutes) compare to the first one to complete its subsequent operation. This finding suggests that individuals' assignments to teams are particularly important for more challenging operations as familiarity gained through these tasks has higher impacts on subsequent productivity and can lead to a significant reduction in the duration of forthcoming operations. One suggestion for hospital managers could be that if for some reason there are individuals who worked together in many challenging tasks in the past, then putting these individuals together in future operations where efficiency is particularly essential (e.g., due to patient condition etc.) would be a good strategy. This is because, through performing challenging tasks together in the past, these individuals have developed a highly efficient team familiarity which would enhance their subsequent productivity.

Finally, teams with high levels of team familiarity seem to be ideal for performing more complex and challenging tasks because the benefits from being familiar with each other increase as the level of task complexity increases. That is, when the hospital faces an exceptionally complex case, it may be a preferable strategy to assign individuals who are highly familiar with each other to this case, as the benefit of team familiarity will be maximized during this complex operation.

As in all empirical studies, our results come with limitations. First, our data comes from a single hospital, so one has to be cautious when interpreting our results. While cardiac surgery operations are an appealing setting with a mix of repetitive and highly critical knowledge-based tasks to study fluid teams, generalizing our results to other fluid team settings requires a careful approach. Moreover, our dataset includes limited information regarding patients' condition before the operation, as well as doctors' characteristics. One would ideally like to have more detailed information about patients' condition and characteristics such as Higgins Score or EuroScore, but unfortunately this was not available from our data source. Similarly, more detailed information on doctors' experience, performance, skill sets, and other characteristics would have been useful to control in predicting team productivity, but these information was unfortunately not available to us. In addition, in this study we focus on productivity implications of team familiarity with the help our highly granular surgery duration data (in minutes), and used in-hospital mortality (a crude measure of quality) as a robustness check to make sure productivity improvements do not come at the expense of increased mortality rates. Future research can investigate how team familiarity can simultaneously affect both productivity and quality in healthcare settings by obtaining more granular quality data such as hospital revisits, clinical and patient reported improvements, among others. Finally, our study joins the stream of work which highlights the importance of team familiarity in fluid teams with a large scale empirical study. As fluid teams become more of a rule than exception in many organizational settings, future research could provide a more in-depth and vivid account of how team familiarity influences team operations and task dynamics with observational or experimental studies.

Despite these limitations our results provide important insights into operations management practice and academic literature. Our study suggests that managers could practice smarter team composition strategies that may allow them to make optimal use of past shared experience of team members, and thus improve productivity of fluid team operations without making use of any additional resources. In addition, in such a critical and costly setting like healthcare, any productivity gain (e.g., reduction in operation time) in surgical operations will not only reduce total costs of treating patients, but it may also translate into better clinical outcomes for patients by reducing infection risks or adverse event probabilities. We also believe that our findings are applicable in other white-collar settings too. In consulting or software teams for example, where a typical project usually lasts months and includes extensive interactions among team members, familiarity dispersion and bottleneck-pairs can also be detrimental for team productivity since relational resources are highly important in those settings as they have been recognized as significant drivers of productivity and efficiency (Staats 2012).

Also, answering the call to decompose team familiarity and further investigate its detailed effects on performance (Huckman et al. 2009), our study introduces two novel concepts into operations management literature, team familiarity dispersion and bottleneck pairs with a focus on their productivity implications. Further disentangling team familiarity by concentrating on how much of the shared work experience is gained during complex tasks in the past; we demonstrate that not all experiences are equivalent, and the higher the ratio of familiarity which is gained from complex tasks, the higher the productivity of the team. Finally, we identify focal task complexity as an important moderator between average team familiarity and productivity, and find that team familiarity is particularly helpful when performing more complex tasks.

3.7 Appendix

Figure 3.2 Distribution of Bottleneck Score

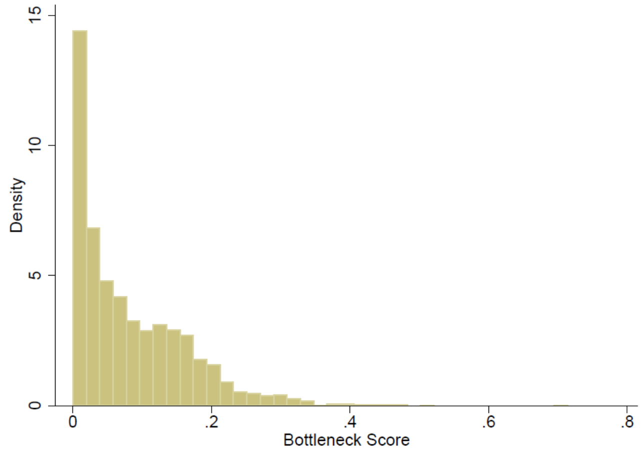


Table 3.3 Regression of Team Familiarity on Surgery Duration using Range for Team Familiarity Dispersion

Variable	Duration	
	Model: (1)	(2)
Team Familiarity Dispersion	0.00003**	0.00003**
	(0.00001)	(0.00001)
Bottleneck-Pair	0.01507**	0.01538**
	(0.00340)	(0.00341)
Severe Familiarity Percentage	-0.40677**	-0.40508**
	(0.11473)	(0.11586)
Average Team Familiarity	-0.03214*	-0.02885*
	(0.01290)	(0.01283)
Average Team Familiarity x Severe		-0.02898*
		(0.01239)
Individual Average Direct Experience	-0.05512**	-0.05604**
	(0.01394)	(0.01392)
Team Size	0.04075**	0.04160**
	(0.00641)	(0.00643)
Severe	0.02479+	0.16449**
	(0.01430)	(0.06001)
Mild	-0.01156+	-0.01121+
	(0.00599)	(0.00599)
Constant	5.81310**	5.30034**
	(0.08224)	(0.08243)
Observations (N)	6129	6129
Adjusted R^2	0.478	0.479
BIC	-1521.933	-1525.260
AIC	-1730.348	-1740.398
Quarter Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes
Procedure Fixed Effect	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes

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

Table 3.4 Regression of Team Familiarity on Surgery Duration using Coefficient of Variation for Team Familiarity Dispersion

Variable	Duration	
	Model: (1)	(2)
Team Familiarity Dispersion	0.4047**	0.3971**
	(0.0750)	(0.0748)
Bottleneck-Pair	0.0167**	0.0170**
	(0.0034)	(0.0034)
Severe Familiarity Percentage	-0.4566**	-0.4538**
	(0.1167)	(0.1176)
Average Team Familiarity	-0.0566**	-0.0529**
	(0.0144)	(0.0143)
Average Team Familiarity x Severe		-0.0270*
		(0.0123)
Individual Average Direct Experience	-0.0249*	-0.0265*
	(0.0112)	(0.0112)
Team Size	0.0512**	0.0518**
	(0.0068)	(0.0069)
Severe	0.0230*	0.1534*
	(0.0098)	(0.0597)
Mild	-0.0112+	-0.0108+
	(0.0060)	(0.0060)
Constant	5.1534**	5.1396**
	(0.0902)	(0.0904)
Observations (N)	6129	6129
Adjusted R^2	0.481	0.482
BIC	-1556.211	-1558.024
AIC	-1764.626	-1773.163
Quarter Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes
Procedure Fixed Effect	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes

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

Table 3.5 Regression of Team Familiarity on Surgery Duration using three dummy Variables for Bottleneck-Pair

Variable	Duration	
	Model: (1)	(2)
Team Familiarity Dispersion	0.0002**	0.0002**
	(0.0000)	(0.0000)
25th Dummy	0.0565**	0.0569**
	(0.0109)	(0.0109)
50th Dummy	0.0198*	0.0210*
	(0.0089)	(0.0089)
75th Dummy	0.0041	0.0046
	(0.0083)	(0.0083)
Severe Familiarity Percentage	-0.4255**	-0.4231**
	(0.1146)	(0.1156)
Average Team Familiarity	-0.0470**	-0.0435**
	(0.0142)	(0.0141)
Average Team Familiarity x Severe		-0.0277*
		(0.0124)
Individual Average Direct Experience	-0.0463**	-0.0474**
	(0.0146)	(0.0145)
Team Size	0.0463**	0.0470**
	(0.0065)	(0.0066)
Severe	0.0242*	0.1576**
	(0.0098)	(0.0600)
Mild	-0.0117+	-0.0113+
	(0.0060)	(0.0060)
Constant	5.3196**	5.3028**
	(0.0822)	(0.0824)
Observations (N)	6129	6129
Adjusted R^2	0.479	0.480
BIC	-1523.859	-1526.135
AIC	-1745.7201	-1754.719
Quarter Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes
Procedure Fixed Effect	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes

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

Table 3.6 Regression of Team Familiarity on Surgery Duration using Distance from Mean for Bottleneck-Pair

Variable	Duration	
	Model: (1)	(2)
Team Familiarity Dispersion	0.0001**	0.0001**
	(0.0000)	(0.0000)
Bottleneck-Pair	0.0002**	0.0002**
	(0.0000)	(0.0000)
Severe Familiarity Percentage	-0.2915**	-0.2891**
	(0.1112)	(0.1122)
Average Team Familiarity	-0.0385**	-0.0347**
	(0.0125)	(0.0125)
Average Team Familiarity x Severe		-0.0275*
		(0.0121)
Individual Average Direct Experience	-0.0310+	-0.0440**
	(0.0159)	(0.0159)
Team Size	0.0462**	0.0469**
	(0.0061)	(0.0062)
Severe	0.0221*	0.1544**
	(0.0097)	(0.0586)
Mild	-0.0112+	-0.0109+
	(0.0059)	(0.0059)
Constant	5.2331**	5.2171**
	(0.0810)	(0.0814)
Observations (N)	6129	6129
Adjusted R^2	0.487	0.488
BIC	-1632.978	-1635.268
AIC	-1841.393	-1851.406
Quarter Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes
Procedure Fixed Effect	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes

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

Table 3.7 Regression of Team Familiarity on Surgery Duration after removing the first 24 Months

Variable	Duration	
	Model: (1)	(2)
Team Familiarity Dispersion	0.0002**	0.0002**
	(0.0000)	(0.0000)
Bottleneck-Pair	0.0161**	0.0165**
	(0.0041)	(0.0041)
Severe Familiarity Percentage	-0.2615*	-0.2590*
	(0.1117)	(0.1122)
Average Team Familiarity	-0.0425*	-0.0396*
	(0.0173)	(0.0173)
Average Team Familiarity x Severe		-0.0234*
		(0.0096)
Individual Average Direct Experience	-0.0390*	-0.0398*
	(0.0179)	(0.0178)
Team Size	0.0418**	0.0425**
	(0.0074)	(0.0074)
Severe	0.0217*	0.1117*
	(0.0108)	(0.0471)
Mild	-0.0082	-0.0077
	(0.0064)	(0.0064)
Constant	5.3632**	5.3471**
	(0.1000)	(0.1003)
Observations (N)	4460	4460
Adjusted R^2	0.466	0.467
BIC	-1430.825	-1433.273
AIC	-1601.766	-1605.823
Quarter Fixed Effect	Yes	Yes
Lead Surgeon Fixed Effect	Yes	Yes
Procedure Fixed Effect	Yes	Yes
Day of the Week Fixed Effect	Yes	Yes

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

Table 3.8 Regression of Team Familiarity on Surgery Duration with Alternative Specification for Task Complexity

Variable	Duration
	Model: (1)
Team Familiarity Dispersion	0.0002**
	(0.0000)
Bottleneck-Pair	0.0179**
	(0.0033)
Team Severe Familiarity Percentage	-0.4082**
	(0.1192)
Average Team Familiarity	-0.0265*
	(0.0128)
Average Team Familiarity x Patient	-0.0112*
	(0.0044)
Individual Average Direct Experience	-0.0479**
	(0.0103)
Team Size	0.0467**
	(0.0046)
Patient	0.0705**
	(0.0221)
Constant	5.170**
	(0.0793)
Observations (N)	6128
Adjusted R^2	0.480
BIC	-1521.1060
AIC	-1742.9680
Quarter Fixed Effect	Yes
Lead Surgeon Fixed Effect	Yes
Procedure Fixed Effect	Yes
Day of the Week Fixed Effect	Yes

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

3.8 References

Aiken, L. S., S. G. West. 1991. *Multiple regression: Testing and interpreting interactions*. Sage, Thousand Oaks, CA.

Altuntas, S., U. Baykal. 2010. Relationship between nurses' organizational trust levels and their organizational citizenship behaviors. *J. of Nursing Scholarship*. **42**(2) 186-194.

Argote, L. 1982. Input uncertainty and organizational coordination in hospital emergency units. *Admin. Sci. Quart*. **27**(3) 420-434.

Avgerinos, E. and B. Gokpinar. 2015. Task Variety in Professional Service Work: When It Helps and When It Hurts. Working Paper, UCL School of Management, UCL, London, UK.

Banker, R. D., J. M. Field, K. K. Sinha. 2001. Work-team implementation and trajectories of manufacturing quality: A longitudinal field study. *Manufacturing & Service Oper. Management.* **3**(1): 25-42.

Berman, S. L., J. Down, C. W. L. Hill. 2002. Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Acad. of Management J.* **45**(1) 13-31.

Bezrukova, K., S. M. B. Thatcher, K. A. Jehn. 2007. Group heterogeneity and faultlines: Comparing alignment and dispersion theories of group composition. Conflict in organizational groups: New directions in theory and practice: 57-92.

Bion, W. R. 1962. *Experience in groups*. Basic Books, NY.

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

Brandon, D. P., A. B. Hollingshead. 2004. Transactive memory systems in organizations: Matching tasks, expertise, and people. *Organ. Sci.***15**(6) 633-644.

Bushe, G. R., A. Chu. 2011. Fluid teams: Solutions to the problems of unstable team membership. *Organ. Dynam.* **40**(3) 181-188.

Campbell, D. J. 1988. Task complexity: A review and analysis. *Acad. of Management Rev.* **13**(1) 40-52.

Catchpole, K. R., A. E. B. Giddings, M. Wilkinson, G. Hirst, T. Dale, M.R. De Leval. 2007. Improving patient safety by identifying latent failures in successful operations. *Surgery*. **142**(1) 102-110.

Chillemi, O., B. Gui. 1997. Team human capital and worker mobility. *J. of Labor Econom*. **15**(4) 567-585.

Choi, J. N., T. Sy. 2010. Group-level organizational citizenship behavior: Effects of demographic faultlines and conflict in small work groups. *J. of Organ. Behav*. **31**(7): 1032-1054.

Cohen, S. G., D. E. Bailey. 1997. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *J. of Management*. **23**(3) 239-290.

Cohen, J., P. Cohen, L. S. Aiken, S. G. West. 2003. *Applied Multiple Regression—Correlation Analysis for the Behavioral Sciences,* 3rd eds. Lawrence Erlbaum, Mahwah, NJ.

Cole, M. S., A. G. Bedeian, R. R. Hirschfeld, B. Vogel. 2011. Dispersion-Composition Models in Multilevel Research A Data-Analytic Framework. *Organ. Res. Methods.* **14**(4): 718-734.

Coleman, J. S. 1988. Social capital in the creation of human capital. *Amer. J. of Sociology***.** 94 95-120.

Colquitt, J.A, J. A. Lepine, C. P. Zapata, R. E. Wild. 2011. Trust in typical and high-reliability contexts: Building and reacting to trust among firefighters. *Academy of Management J.* **54**(5) 999-1015.

Cook, L. M. 2013. Can nurses trust nurses in recovery reentering the workplace? *Nursing* **43**(3) 21-24.

Cramton, C. D. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. *Organ. Sci.* **12**(3) 346-371.

Cross, R., A. Parker, L. Prusak, S. P. Borgatti. 2001. Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organ. Dynam*. **30**(2) 100-120.

Davenport, D. L., W. G. Henderson, C. L. Mosca, S. F. Khuri, R. M. Mentzer. 2007. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J. of the Amer. College of Surgeons*. **205**(6) 778-784.

Dawson, J. F., A. W. Richter. 2006. Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. *J. of Appl. Psych.* **91**(4) 917-926.

De Jong, B. A., K. T. Dirks. 2012. Beyond shared perceptions of trust and monitoring in teams: implications of asymmetry and dissensus. *J. of Appl. Psych*. **97**(2): 391.

Dineen, B. R., R. A. Noe, J. D. Shaw, M. K. Duffy, C. Wiethoff. 2007. Level and dispersion of satisfaction in teams: Using foci and social context to explain the satisfaction-absenteeism relationship. *Acad. of Management J*. **50**(3) 623-643.

Drucker, P. F. 1999. Knowledge-worker productivity: The biggest challenge. *California Management Rev*. **41**(2) 79-94.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Admin. Sci. Quart.* **44**(2) 350-383.

Edmondson, A. C., R. M. Bohmer, G. P. Pisano. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Admin. Sci. Quart*. **46**(4) 685-716.

Edmondson, A. C., I. M. Nembhard. 2009. Product Development and Learning in Project Teams: The Challenges Are the Benefits. *J. of Product Innovation Management*. **26**(2) 123-138.

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*. **34**(2) 197-224.

Espinosa, J. A., S. A. Slaughter, R. E. Kraut, J. D. Herbsleb. 2007. Familiarity, complexity, and team performance in geographically distributed software development. *Organ. Sci*. **18**(4) 613-630.

Evans, C. R., K. L. Dion. 1991. Group cohesion and performance a meta-analysis. *Small Group Res.* **22**(2) 175-186.

Faraj, S., L. Sproull. 2000. Coordinating expertise in software development teams. *Management Sci.* **46**(12) 1554-1568.

Feldman, D. C. 1984. The development and enforcement of group norms. *Acad. of Management Rev*. **9**(1) 47-53.

Field, J. M., K. K. Sinha. 2005. Applying process knowledge for yield variation reduction: A longitudinal field study. Decision Sciences. **36**(1) 159-186.

Fleming, M., S. Smith, J. Slaunwhite, J. Sullivan. 2006. Investigating interpersonal competencies of cardiac surgery teams. *Canadian J. of Surgery*. **49**(1) 22.

Galbraith, J. R. 1973. *Designing Complex Organizations*. Addison- Wesley Longman Publishing Co, Inc., Boston, MA.

Gardner, H. K., F. Gino, BR. Staats 2012. Dynamically integrating knowledge in teams: Transforming resources into performance. *Acad. of Management J.* **55**(4) 998-1022.

Gawande, A. 2010. Checklist Manifesto, Vol. 200. New York: Metropolitan Books.

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.

Gersick, C. J., J. R. Hackman. 1990. Habitual routines in task-performing groups. *Organ. Behav. and Human Decision Processes*. **47**(1) 65-97.

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.

Gittel, J. H. 2002. Coordinating mechanisms in care provider groups: Relational coordination as a mediator and input uncertainty as a moderator of performance effects. *Management Sci*. **48**(11) 1408-1426.

Goldratt, E. M., J. Cox. 1984. *The Goal: A Process of Ongoing Improvement.* North River Press, Croton-on-Hudson, NY.

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

Gruenfeld, D. H., E. A. Mannix, K. Y. Williams, M. A. Neale. 1996. Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organ. Behav. and Human Decision Processes* **67**(1) 1-15.

Gruenfeld, D. H., P. V. Martorana, E. T. Fan. 2000. What do groups learn from their worldiest members? Direct and indirect influence in dynamic teams. *Organ. Behav. and Human Decision Processes* **82**(1) 45-59.

Hackman, J. R. 1987. The design of work teams. J. W. Lorsch eds. *Handbook of Organizational Behavior*. Prentice Hall, Englewood Cliffs, NJ, 315 342.

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

Harrison, D. A., K. J. Klein. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Acad. of Management Rev*. **32**(4) 1199-1228.

Harrison, D. A., S. Mohammed, J. E. McGrath, A. T. Florey, S. W. Vanderstoep. 2003. Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psych*. **56**(3) 633-669.

He, B., F. Dexter, A. Macario, S. Zenios. 2012. The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Oper. Management*. **14**(1) 99-114.

Hoffmann, D. A., Lei, Z., & Grant, A. M. 2009. Seeking help in the shadow of a doubt: The sensemaking processes underlying how nurses decide whom to ask for advice. *J. of Appl. Psych.* **94**(5): 1261-1274.

Hollingshead, A. B. 2001. Cognitive interdependence and convergent expectations in transactive memory. *J. of Personality and Soc. Psych.* **81**(6) 1080.

Hornsey, M. J., M. A. Hogg. 2000. Assimilation and diversity: An integrative model of subgroup relations. *Personality and Social Psych. Rev*. **4**(2) 143-156.

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

Hopp, W. J., M. L. Spearman. 2000. *Factory Physics: Foundations of Manufacturing Management,* 2nd ed. McGraw-Hill, Burr Ridge, IL.

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 Oper. Management*. **13**(3) 310-328.

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

Jackson, S. E., A. Joshi, N. L. Erhardt. 2003. Recent research on team and organizational diversity: SWOT analysis and implications. *J. of Management*. **29**(6) 801-830.

Kane, A. A., L. Argote, J. M. Levine. 2005. Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organ. Behav. and Human Decision Processes* **96**(1) 56-71.

Katz, R. 1982. The effects of group longevity on project communication and performance. *Admin. Sci. Quart.* **27**(1) 81-104.

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

KC, D. S., C. Terwiesch. 2011. The effects of focus on performance: Evidence from California hospitals. *Management Sci.* **57**(11) 1897-1912.

KC, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Oper. Management*. **14**(1) 50-65.

Klein, K. J., S. W. Kozlowski. 2000. Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions, Jossey-Bass.

Lee, D. K. K., S. A. Zenios. 2009. Optimal capacity overbooking for the regular treatment of chronic conditions. *Oper. Res.* **57**(4) 852-865.

Levine, J. M., R. L. Moreland. 1999. Knowledge transmission in work groups: Helping newcomers to succeed. J. M. Levine, D. M. Messick, L. L. Thomson eds. *Shared Cognition in Organizations: The Management of Knowledge,* Psychology Press, NY.

Lewis, K. 2003. Measuring transactive memory systems in the field: Scale development and validation. *J. Appl. Psych.* **88**(4) 587–604

Li, J., D. C. Hambrick. 2005. Factional groups: A new vantage on demographic faultlines, conflict, and disintegration in work teams. *Acad. of Management J*. **48**(5) 794-813.

Liao, H., D. Liu, R. Loi. 2010. Looking at both sides of the social exchange coin: A social cognitive perspective on the joint effects of relationship quality and differentiation on creativity. *Acad. of Management J*. **53**(5) 1090-1109.

Lingard, L., G. Regehr, B. Orser, R. Reznick, G. R. Baker, D. Doran,S. Espin,J. Bohnen, S. Whyte. 2008. Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Arch. of Surgery.* **143**(1) 12.

Makary, M. A., J B. Sexton, J. A. Freischlag, C. G. Holzmueller, E. A. Millman, L. Rowen, P. J Pronovost. 2006. Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. *J. of the Amer. College of Surgeons*. **202**(5) 746-752.

Moreland, R. L. 1999. Transactive memory: Learning who knows what in work groups and organizations. L. Thomson, D. Messick, J. Levine, eds. *Sharing Knowledge in Organization.* Lawrence Erlbaum, Hillsdale, NJ.

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. *J. of Organ. Behav*. **27**(7) 941-966.

Nemeth, C. J. 1986. Differential contributions of majority and minority influence. Psych. Rev. **93(1)** 23-32.

Paletz, S. B. F., C. D. Schunn. 2011. Assessing group-level participation in fluid teams: Testing a new metric. *Behav. Res. Methods*. **43**(2) 522-536.

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*. **47**(6) 752-768.

Powell, A., S. Savin, N. Savva. 2012. Physician Workload and Hospital Reimbursement: Overworked Physicians Generate Less Revenue Per Patient. *Manufacturing & Service Oper. Management*. **14**(4) 512-528.

Pullon S., Competence, respect and trust: Key features for successful interprofessional nurse-doctor relationships. 2008. *J. of Interprofessional Care.* **22**(2) 133-147.

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*. **51**(6) 869-881.

Schroder, H., M. Driver, S. Streuferi. 1967. *Human information processing*. Holt, Rinehart and Winston, New York.

Siemsen, E., A.V. Roth, S. Balasubramanian. 2008. How motivation, opportunity, and ability drive knowledge sharing: The constraining-factor model. *J. of Oper. Management*. **26**(3) 426-445.

Staats, BR. 2012. Unpacking Team Familiarity: The Effects of Geographic Location and Hierarchical Role. *Production and Oper. Management*. **21**(3) 619-635.

Summers, J. K., S. E. Humphrey, G. R. Ferris. 2012. Team member change, flux in coordination, and performance: Effects of strategic core roles, information transfer, and cognitive ability. *Acad. of Management J*. **55**(2) 314-338.

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

Tushman, M. L., D. A. Nadler. 1978. Information Processing as an Integrating Concept in Organizational Design. *Acad. of Management Rev*. **3**(3) 613-624.

Van Knippenberg, D., C. K. W. De Dreu, A. C. Homan. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology* **89**(6) 1008–1022.

Van Knippenberg, D., M. C. Schippers. (2007). Work group diversity. *Annual Review of Psychology* **58** 515–541.

Van de Ven, A. H., A. L. Delbecq, R. Koenig Jr. 1976. Determinants of coordination modes within organizations. *Amer. Sociological Rev*. **41**(2) 322-338.

Wegner, D. M. 1986. Transactive memory: A contemporary analysis of the group mind. G. Mullen, G. Goethals, eds. *Theories of Group Behavior*. Springer-Verlag, New York, 185-208.

Wegner, D. M., T. Giuliano, P. T. Hertel. 1985. Cognitive interdependence in close relationships. W. J. Ickes, eds. *Compatible and Incompatible Relationships,* Springer-Verlag, New York. 253-276.

Weick, K. E., K. H. Roberts. 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Admin. Sci. Quart.* **38**(3) 357-381.

Chapter four – The Role of Non-Clinical Workforce on Efficiency and Quality of Patient Service: Evidence from NHS Medical Helpline

Healthcare organizations increasingly rely on a mix of clinical and non-clinical health personnel in providing innovative healthcare services such as medical helplines. While these can offer significant cost and patient access advantages, determining the right mix of health personnel is a major challenge in such settings. In this study, making use of a dataset (i.e., England's National Health Service (NHS)'s new 111 non-emergency helpline), we investigate the effect of non-clinical labor mix on efficiency and quality of patient service. Our results indicate that while non-clinical workforce increases the efficiency of patient service by reducing abandoned calls, it may lead to new inefficiencies through misuse of critical resources (i.e., unnecessary ambulance dispatches) and it may reduce the quality outcome of the patient service.

4.1 Introduction

Healthcare organizations are facing an increasing pressure from both public and private payers to reduce their operational expenses while retaining high levels of service efficiency, quality and access to patients. This ever-growing challenge has highlighted the need for the development of innovative services such as telephone helplines that provide easy access to medical support to patients at a low cost. In providing such innovative services, healthcare organizations rely on a mix of clinical and non-clinical health personnel. However, as highlighted in the World Health Report titled: "Health Systems: Improving Performance" (World Health Organization 2000), determining the right mix of health personnel is a major challenge for most healthcare organizations and health systems (Buchan and Dal Poz 2002). With this challenge in mind, our study aims to investigate efficiency and quality implications of non-clinical labor mix in delivering new and innovative forms of health services. More specifically, we focus on telephone based health services which have become popular recently due to their cost and access advantages and have been adopted by a wide number of public and private healthcare organizations (Spiegelman 2000, Rhian and Claudio 2003).

One such innovative service is NHS 111 telephone helpline, which was introduced in 2010 by National Health Service (NHS) in England to provide easy and low cost patient access to medical help or advice in "urgent but not life-threatening" situations (life-threatening emergency line is 999 in England). This is a 24 hours a day, 365 days a year, free to use service. People call this number when they need medical help fast, but it's not a 999 emergency; when they don't know who to call for medical help; when they think they need to go to an emergency department but they are not sure; or when they require health information or advice about what to do next (NHS 2014). The backbone of NHS 111 operations is a mix of non-clinical and clinical personnel. The first group - “call handlers” are the first point of contact, they are not medically trained, and they make an initial assessment of callers by using special software. The second group involves medically trained “clinical advisors” who provide support when a call needs further assessment or health advice. By making use of this unique setting and dataset which enables us to investigate the effect of non-clinical labor mix on service efficiency and quality of healthcare organizations, our study extends existing literature in healthcare operations management on four major fronts.

First, while there has been considerable amount of work in healthcare operations management literature investigating the implications of staffing levels and workforce management in hospitals (Li and Benton 2006, Chow et al. 2011, He et al. 2012, Green et al. 2013), little is known about the effect of non-clinical staff members on service efficiency and quality of healthcare organizations. At the same time such organizations adopt both role substitution and delegation (Dubois and Singh 2009). Role substitution involves using non-professionally qualified workers to substitute more expensive qualified ones while role delegation includes breaking down job demarcations so that parts of the simplest tasks can be performed by less qualified and lower-cost workers. While both role substitution and delegation are used in order to decrease overall cost (Department of Health 2000), evidence of their impact is limited and conflicting (Dubois and Singh 2009). Several studies have pointed out that these two practices may reduce workforce cost but at the same time they can lead to decreased quality increasing therefore overall cost for healthcare organizations (Powers et al. 1990, Garfink et al. 1991, Bostrom and Zimmerman 1993). Our paper contributes to this stream of literature by focusing on these practices used by medical helplines, a highly important but understudied component of healthcare operations and examining their impact on both cost and clinical outcome.

Second, while patient is at the heart of any healthcare operation, most of the existing studies in healthcare operations management focus on either system performance measures such as capacity utilization (Salzarulo et al. 2011, Cayirli et al. 2008), average waiting times (Cayirli et al. 2012), and operational failures (Tucker and Spear 2006); or system level health outcomes such average mortality rates (KC and Staats 2012), likelihood of multiple surgeries (Anderson et al. 2014) or length of stay for the patient (Andritsos and Tang 2014). In this study, we concentrate on patient service and simultaneously investigate two dimensions of performance with a nuanced approach. We first examine the impact of workforce management decisions regarding the level of non-clinical and clinical staff on patient service efficiency by looking at (i) number of abandoned calls and (ii) use of costly ambulance dispatches. Next, focusing on the service quality experienced by the primary end user – the patient, we investigate how non-clinical labor mix affect the quality outcome of the call. As the critical challenge for many healthcare systems is to achieve multiple objectives at the same time, we think it is important to consider multiple performance dimensions when evaluating any health management decision. Consequently, considering patient at the center of healthcare operations, our study complements existing studies by disentangling efficiency and quality outcomes when healthcare organizations rely on workers with mixed qualifications.

Thirdly, with our focus on NHS 111 which is effectively a call center for medical needs, our study is also related to a large stream of research in operations management on call center workforce management (Aksin et al. 2007, Bhulai et al. 2008, Aksin et al. 2015) and recent work on staff mixing and labor flexibility decisions in service operations (Kuo et al. 2014, Kesavan et al. 2015). Our empirical work contributes to this stream of literature by examining workforce mix decisions in a novel healthcare setting.

Finally, our results are significant both statistically and economically. Our models indicate that non-clinical labor mix can significantly decrease abandoned calls while increasing ambulance dispatches and patients who experience worse outcomes after using the medical helpline. These findings suggest that efficient workforce allocation can have an important and meaningful impact on efficiency, cost and quality of service for healthcare organizations.

In the following section we describe our dataset and setting. Next, we develop our hypotheses and define our variables and empirical strategy. We then present our results and our robustness checks. Finally, we discuss our findings and conclusions from this study.

4.2. Data and Setting

Our setting is the 111 medical helpline in the UK. NHS advises patients to call 111 when a general practitioner (GP) is not an option, the patient cannot wait due to her condition or she needs guidance on what to do next (NHS 2012). Through this new medical helpline NHS also aims to reduce its cost by phasing out the £123 million-a-year NHS Direct line and directing patients away from accident and emergency departments (MailOnLine 2013). According to NHS 25% of visits to accident and emergency departments could have been treated elsewhere in their community or even self-treated. In addition, 33.8% of calls that the ambulance service receives in England are classified as urgent rather than emergency (NHS 2011). Between August 2010 and

November 2015 NHS 111 received 32,941,880 calls and has dispatched 2,878,564 ambulances. It is currently stated that around one million patients per month use NHS 111 (NHS 2014).

The system operates as follows: First the call is answered by a call handler who uses a special clinical assessment software and provides health advice to the caller, refers her to another service, dispatches an ambulance or triages her (NHS 2012). Then, if necessary the patient is transferred to a clinical advisor (usually a nurse or paramedic). There are two ways for this transfer to occur: Either the caller is live or on hold (i.e., warm transfer) or the call is ended and the caller is offered a call back. Finally, the clinical advisor will give the caller healthcare advice, recommend her to attend a specific service or dispatch an ambulance.

Call handlers have no medical background nor experience and receive 6 weeks of training in using the software tool and basic health and safety (Anderson and Roland 2015) (although it has been reported that the training can last four-instead of six-weeks (The Telegraph 2015)) including nine days in classroom, written assessments (NHS 2010) and lengthy periods of monitoring and review of calls from clinical advisors (Anderson and Roland 2015). Clinical advisors are typically individuals with medical background such as nurses or paramedics and advise a patient if the call handler transfers the call to them. The ratio of clinical advisors to call handlers varies typically from one to four to one to six (Anderson and Roland 2015).

Our dataset involves monthly NHS 111 call data from different regions (i.e., call centers) across England between August 2010 to October 2015 and includes monthly information regarding the number of calls received, answered, abandoned by the caller, transferred to a clinical advisor, given advice, recommended to attend a specific service or resulted in an ambulance dispatch. In addition, we have data regarding average episode length, transfer waiting time and time worked by call handlers and clinical advisors respectively. Finally, our dataset includes semi-annual data regarding whether the patient's problem was solved or got worse. Because NHS 111 is a new service and it started in different regions gradually over time, we have full data for some regions and partial or very limited data in other regions. After removing those observations with missing data, our final dataset includes an unbalanced panel of 642 observations from 28 different call centers. In addition, for our third hypothesis, we use 459 observations from 23 regions due to missing data regarding patient experience.

4.3 Hypotheses Development

In developing and testing our hypotheses, we define non-clinical labor mix as the ratio of time worked by call handlers to time worked by clinical advisors. Similar to Kesavan et al. (2015), we divide total time worked by call handlers by the total time worked by clinical advisors in order to enable comparison across different regions. Non-clinical employees such as call handlers are less costly for the organization and they are more readily available after providing them limited training. While having a higher non-clinical labor mix can be helpful for efficiency of patient service, this may not be the case for quality. Clinical workforce is clearly more expensive to hire and maintain from a smaller pool of qualified healthcare personnel, but they also provide a higher quality service to patients. Considering these trade-offs associated with labor mix in healthcare settings, we develop and test three hypotheses regarding efficiency and quality of patient service.

In our first hypothesis, we focus on patient service efficiency by using the number of abandoned calls. Specifically, we expect to observe a negative relationship between non-clinical labor mix and number of calls abandoned by the caller while she is on queue or while talking to a call handler. When a patient calls an inbound call center a call handler will answer that call. However, if there is no available call handler to immediately answer that call the patient will be put on hold and placed in a queue. In that case the patient may abandon her/his call by hanging up either right after she/he is put on hold or after waiting for some time in queue without receiving any service. Higher numbers of agents that can respond to a call have been associated with lower abandoned rates from the customers in the call center literature (Saltzman 2005, Saltzman and Mehrotra 2007, Armony et al. 2007, Aksin et al. 2007).

Similarly, for healthcare medical helplines increasing the number of available agents leads to a lower number of abandoned calls (Gustafson 1999). Because in our setting call handlers are the ones initially answering calls to 111, we expect to observe a decrease in the number of abandoned calls as the number of non-clinical call handlers increases. On the contrary, the number of abandoned calls will not be affected by the number of available clinicians since they are not the ones answering the call at the first stage. In addition, while talking to a call handler the call can be terminated by either side. We therefore believe that if the number of call handlers is low, the latter ones will be more willing to terminate the call if the patient asks them to do so. We therefore expect that:

Hypothesis 1: *Non-clinical labor mix has a negative relationship with the number of abandoned calls.*

We next focus on efficiency in terms of the resources used. Because ambulance dispatches are the most expensive form of resource provided to patients in urgent medical assistance, their unnecessary use will result in major waste of highly critical resources for NHS (Turner et al. 2006). In addition, it has been reported that NHS 111 has been sending ambulances even for minor cases such as cut fingers (The Telegraph 2015), increasing therefore overall cost for NHS. Consequently, we examine the relationship between non-clinical labor mix and ambulance dispatches. Specifically, we expect to observe a positive relationship between non-clinical labor and ambulance dispatches for three reasons.

First, because call handlers are not medically trained, they are more likely to misjudge the criticality of a situation with compare to clinically trained advisers who can assess the situation more easily. As a result, when they face a non-trivial situation, they are more likely to dispatch an ambulance even if it is not necessary and other less costly actions such as referring them to a regular GP appointment would have been sufficient.

Second, call handlers will be quite risk averse and certainly more than clinical advisors. Call handlers typically have no specialized experience and receive a few weeks' training before start working at NHS 111 (Anderson and Roland 2015, The Telegraph 2015). They are therefore low trained individuals with very limited medical knowledge handling a wide range of situations from common colds to heart attacks (The Telegraph 2015). As a result, we believe that they will be quite risk averse since past studies have shown that low confidence and knowledge make people unwilling to take risks (Krueger et al. 1994, Campbell et al. 2004). Hence we expect them to dispatch more easily an ambulance than a clinical advisor if they feel that they do not assess the situation of the caller correctly. As a senior trainer of call handlers in NHS 111 admitted: "It may not be a particularly management orientated view, but I would rather we get an ambulance out to somebody that doesn't need it than not get an ambulance out to somebody that does need it" (The Telegraph 2015).

Third, a high ratio of call handlers to clinical advisors implies that clinical advisors will not be easily available to receive a transferred call. In fact, it has been reported that in many cases there is no clinical advisor present at the call center, only call handlers (MailOnLine 2015). As a result, when the call handler faces a situation where she is not sure whether to dispatch an ambulance right away or transfer the call to a clinical advisor for further assessment, she is likely to choose the former if she thinks that transferring the call will require a long waiting time for the patient due to limited clinical advisor availability. We therefore predict that:

Hypothesis 2: *Non-clinical labor mix has a positive relationship with the number of ambulance dispatches.*

Next, focusing on the service quality experienced by the patient, we investigate how the ratio of call handlers to clinical advisors affects the quality outcome of the call. Specifically, we make use of a patient survey conducted by call centers where patients report if their problem was resolved or not and examine the effect of non-clinical workers on the number of problems that got worse despite using the medical helpline.

As we also discuss in the introduction using non-clinical labor workforce in order to replace high-qualified employees (i.e., role substitution) or perform the simple parts of the job (i.e., role delegation) has been associated with worse clinical outcomes for the patients (Powers et al. 1990, Garfink et al. 1991, Bostrom and Zimmerman 1993). Substituting less qualified personnel for highly qualified ones will inevitably lead to lower quality care for healthcare organizations since higher ratios of qualified workers have been associated with better outcomes and fewer adverse events for patients (Duboi and Singh 2009).

Similarly, in our setting increasing the number of call handlers (especially at the expense of clinically trained personnel) will lead to lower quality care for the patients. Since non-clinical labor (call handlers) is not as qualified and medically trained as clinical advisors, we don’t expect it to be as effective as clinical handlers in addressing patients’ problem. Therefore, the higher the non-clinical labor mix, the more patients’ problems will become worse.

In addition, we believe that call handlers will have issues in persuading patients to follow their advice. Patients refusing to follow treatments and recommendations constitutes a pressing global problem for healthcare organizations and medical staff (Bang and Ragnemalm 2012). Competent patients can make irrational decisions regarding their health (Brock and Wartman 1990) due to several reasons that include-among others-distrust for the medical staff and problems of communication (Connelly and Campbell 1987). Swindell et al. (2010) present a number of cognitive reasons that can lead patients to disregard a recommendation causing increased costs for the organizations and worse clinical outcomes for the patients (Appelbaum and Roth 1983).

Similarly, in our setting we believe that patients will not comply always with the recommendation they receive especially when they receive it from a call handler and not from a clinical advisor. Past studies have shown that most individuals consider the source when receiving a persuasive message (Hovland et al. 1953, Petty and Cacioppo 1986, Briñol and Petty 2009, Smith et al. 2012) regarding the source more credible when it is perceived as expert (Hovland and Weiss, 1951, McGinnies and Ward 1980, Gottlieb and Sarel, 1991, Clark et al. 2012, Smith et al. 2012). Expert sources are expected to present more valid information and suggestions compared to non-expert ones (Clark et al. 2012) and are therefore more likely to persuade the receiver of their suggestions (Ziegler et al. 2002, Clark and Evans 2014), especially in situations in which the receiver has low processing ability (Petty et al. 1981, Clark et al. 2012) or confidence (Tormala et al. 2006). We therefore expect that patients will follow more likely a recommendation when they receive it from a source they consider expert since their medical knowledge and confidence are both low. And since clinical advisors are considered to be more credible than call handlers we believe that call handlers will have more issues in persuading patients to follow their recommendations, which will lead to worse clinical outcomes with compare to the ones of clinical advisors. In addition, since the service is a medical helpline it is likely that call handlers with limited medical training will experience communication problems with the patients leading to the rejection of their recommendation to them. Hence, we expect that:

Hypothesis 3: *Non-clinical labor mix has a negative effect on the quality outcome of the call.*

4.4 Variables and Empirical Strategy

4.4.1 Dependent Variables.

Abandoned Calls. For our first hypothesis we measure service efficiency using the number of abandoned calls of call center i at month t. These calls represent how many times a caller chose to hang up while either waiting in the queue or talking to a call handler and how many times a call handler terminated the call too. Finally, we divide this variable by the total number of calls offered of each call center every month in order to enable comparison across all call centers.

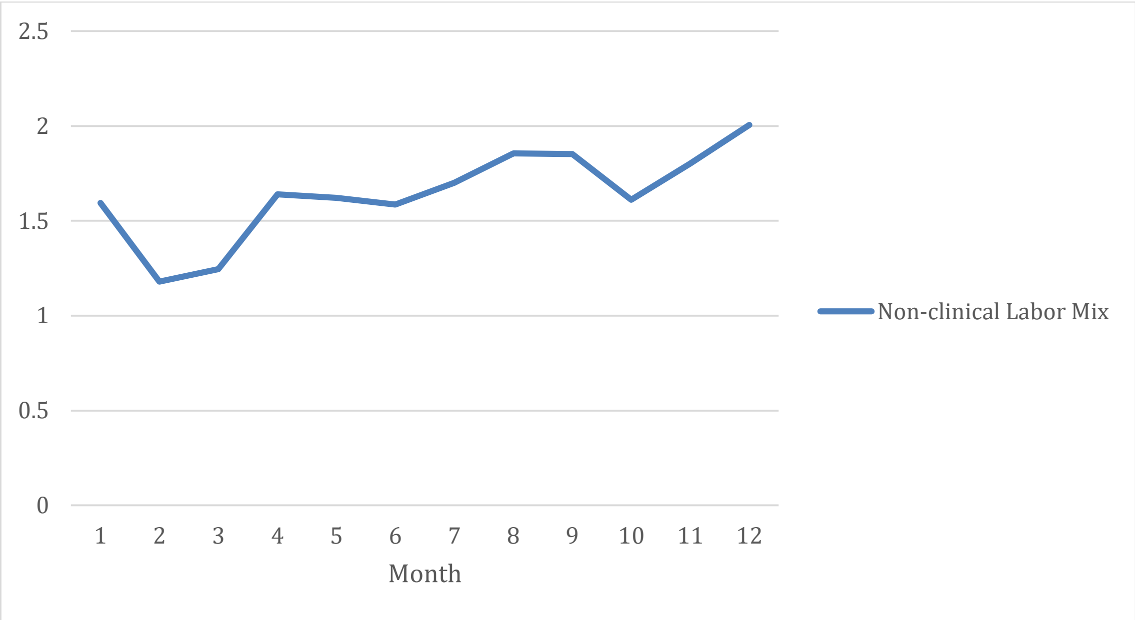
Ambulance Dispatches. For our second hypothesis we use the number of ambulance dispatches of call center i during month t in order to capture the efficiency of the service in terms of resource usage. Due to their high cost ambulance dispatches constitute an indicator of expensive resources used and therefore overall cost. We also divide this variable by the total number of calls offered of each call center every month.

Worse Problems. For our third hypothesis we use the percentage of patients that reported that their problem got worse after using the medical helpline of center i at month t. This way we examine the effect of non-clinical labor-mix on the quality of the service in terms of medical outcome. This variable is reported in our dataset every six months (and not every month like all the other variables). In order to deal with that for our main model we use the same value of this variable for every month within the six months of the survey. We investigate this approach further in the robustness checks section.

4.4.2 Independent Variable

Non-Clinical Labor Mix. We define our independent variable by dividing the number of hours worked by call handlers by the number of hours worked by clinical advisors. We normalize each variable by the hours of clinical workers to enable comparison across all call centers. This variable is defined similar to Kesavan et al. (2015) and captures the ratio of non-clinical workers to clinical ones of the call center i at month t. Figure 4.1 shows the mix of a typical call center during a year. As one can notice there is significant variation of the staffing level during the year.

Figure 4.1 Non-clinical Labor Mix



4.4.3 Control Variables

Offered Calls. In spite of controlling for the call center fixed effect we also control for the number of offered calls through NHS 111 of center i at month t. Larger regions receive more calls resulting in more abandoned calls and ambulance dispatches.

Answered Calls. This variable captures the percentage of answered calls for each center. We divide the number of answered calls of center i at month t by the number of offered calls in order to control for each different region.

Transferred Calls. As we also mention above this variable captures the percentage of calls that were transferred to a clinically trained advisor. We divide the number of transferred calls of center i at month t by the number of offered calls in order to control for each different region.

Transferred Waiting Time. This variable indicates the total time (in minutes) each patient had to wait on average in order to be transferred to a clinically trained advisor of center i during month t.

Episode Length. This variable indicates the total time (in minutes) an episode lasted on average (since the beginning of the call until the end of it) of center i at month t. Finally, we control for center-fixed effect to account for region-specific factors, year and quarter-fixed effects to account for time-specific factors and trends. We use a panel fixed effect model with standard errors clustered by call center and dummy variables indicating the year and the quarter. Figures 4.2, 4.3, and 4.4 provide information of the basic monthly variables of a typical call center during a typical year. As one can notice there is also important variation in most of these variables during a year.

Figure 4.2 Basic Variables per Year

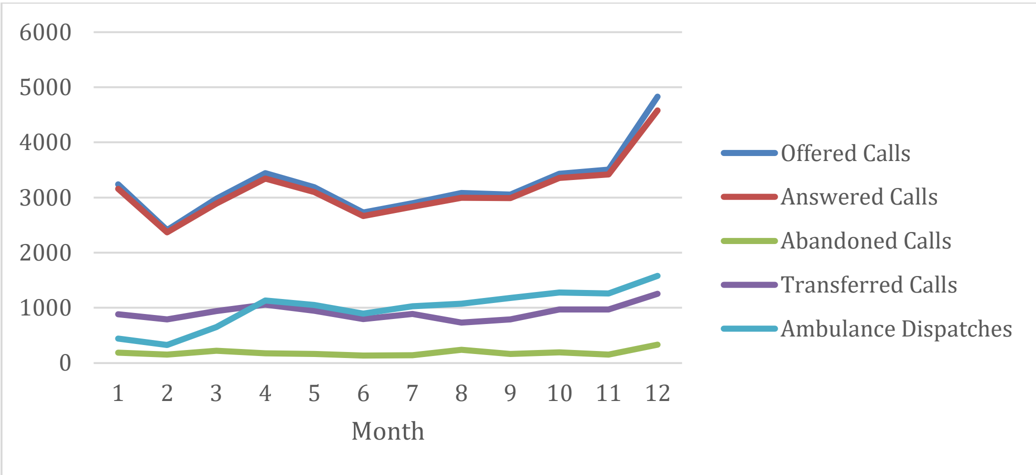


Figure 4.3 Percentages of Basic Variables

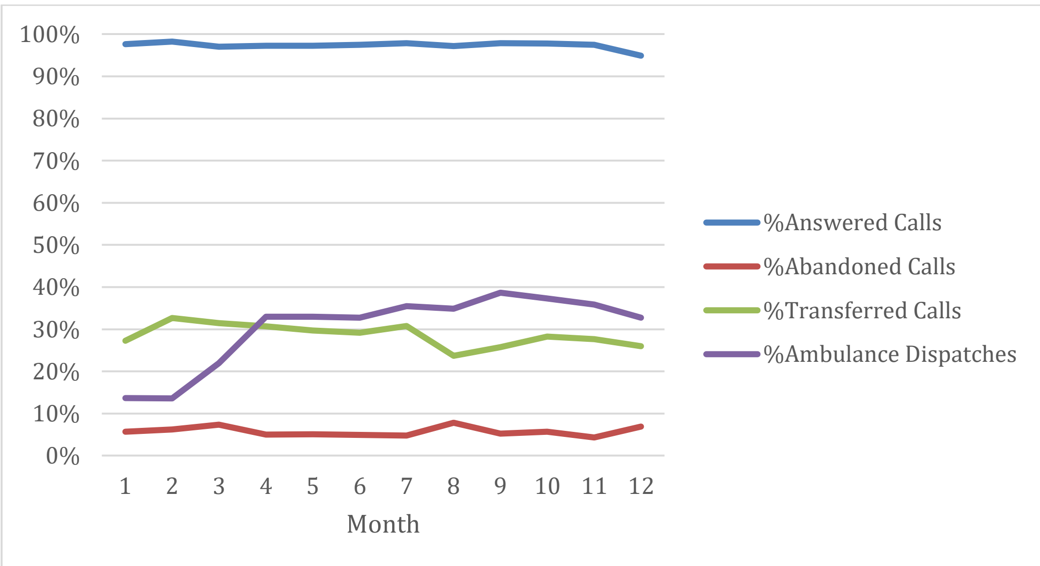
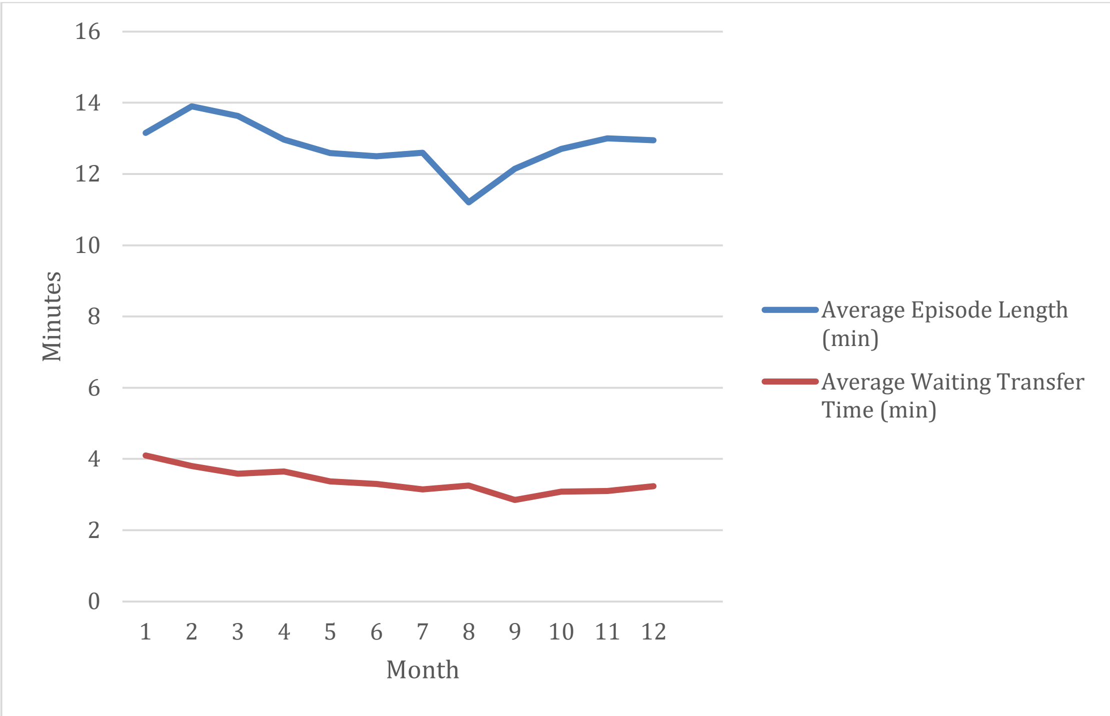


Figure 4.4 Episode Length and Transfer Time



4.4.4 Model Specification

Summary statistics and correlations for our variables are presented at Table 4.1. Some of our variables are normalized as we also explained above. Table 4.1 reports the values for the normalized variables. Our models for our three hypotheses are the following:

$$Abandoned\ Calls_{it} = a_i + a_1(Non\text{-}clinical\ labor\ mix_{it}) + a_2 controls_{it} + \varepsilon_{it}$$

$$Ambulance\ Dispatches_{it} = \beta_i + \beta_1(Non\text{-}clinical\ labor\ mix_{it}) + \beta_2 controls_{it} + \theta_{it}$$

$$Worse\ Problems_{it} = \gamma_i + \gamma_1(Non\text{-}clinical\ labor\ mix_{it}) + \gamma_2 controls_{it} + \varphi_{it}$$

4.4.5 Endogeneity Issues

An important issue is that of endogeneity. In order to deal with that, we use an instrumental variable approach in our main models. For our hypotheses we instrument Non-clinical labor mix and use the lagged Non-clinical labor mix. Specifically, we use the labor mix of each call center of the previous month. It is a widely used instrument when it comes to labor-related studies (Bloom and Van Reenen 2007, Siebert and Zubanov 2010, Tan and Netessin 2014, Kesavan et al. 2015) and it is a valid instrument for our study too: While, it does not affect the performance of the current month it does affect the labor-mix of it. It is not easy for the call center management to significantly change the labor-mix from month to month. We therefore expect our instrument to be

highly correlated with the current month's Non-clinical labor mix (which is also confirmed by the high values of R-squared in the first stage of our three hypotheses).
Finally, for all our hypotheses we check the validity of our instruments. First, we check the values of R-squared of the regression of the first stage. For our first two hypotheses R-squared is equal to 0.793 and our instrument is significant at 1% with a positive coefficient. For our third hypothesis R-squared is equal to 0.241 (the change with compare to H1, H2 and H3 is caused by the different number of observations) and our instrument is significant at 1% with a positive coefficient. Next, we check the F-statistics of the excluded instruments from the first stage and find that they are all well above 10 indicating that they are not weak according to Staiger and Stock (1997). We therefore believe that our instrument is valid.

4.5 Results

Table 4.2 shows the results for all our hypotheses. At model 1, 3, and 5 we use Ordinary Least Square analysis without correcting for endogeneity. In all our models the coefficients of Non-clinical labor mix provide full support for all our hypotheses. Nonetheless, since these coefficients may be biased because of endogeneity we discuss the coefficients obtained after controlling for it.
Model 2 shows our results for our first hypothesis. The coefficient of Non-clinical labor mix is significant at 1% and negative providing support for our first hypothesis. An increase of 50% of this variable (from 4:1 to 6:1) decreases abandoned calls on average by 6.6%. Remember that we divide our abandoned calls by the number of offered calls so essentially 6.6% is the decrease in the percentage of abandoned calls hence the actual number of abandoned calls will be quite high: According to our dataset this is translated in around 75 less abandoned calls for every call center per month.
Model 4 shows our result for our second hypothesis. The coefficient of Non-clinical labor mix is significant at 5% and positive providing support for our second hypothesis. An increase of 50% of this variable increases ambulance dispatches on average by 0.74% Similarly, this increase refers to the percentages of calls that resulted in an ambulance dispatch. Specifically, this is translated in around 15 more ambulance dispatches for every call center per month.
Model 6 shows our result for our third hypothesis. The coefficient of Non-clinical labor mix is significant at 1% and positive providing support for our third hypothesis. An increase of 50% of this variable increases on average worse problems by 35.93%. Again this dependent variable captures the percentage of patients that reported that

Table 4.1 Summary Statistics

Variable	Mean	Std. Dev.	Min	Max	1	2	3	4	5	6	7	8	9
1.Abandoned Calls	0.045	0.045	0	0.250	1								
2. Ambulance Dispatches	0.081	0.017	0.033	0.133	0.154**	1							
3. Worse Problems	0.059	0.050	0	0.239	0.088+	-0.132**	1						
4. Non-Clinical Labor Mix	3.131	3.582	0.320	59.083	0.069+	0.083*	0.064	1					
5. Offered Calls	25,342	25,993	2,410	153,497	0.071+	0.013	-0.328**	-0.081*	1				
6. Answered Calls	0.927	0.082	0.582	1	-0.231**	0.413**	0.112*	-0.100+	0.066+	1			
7.Transferred Calls	0.220	0.061	0.094	0.422	-0.396**	0.073+	0.113*	0.000	-0.078*	0.506**	1		
8.Transfer Waiting Time	1.284	1.448	0	7.233	0.036	-0.038	0.058	-0.036	-0.189**	0.231**	0.153**	1	
9.Episode Length	13.347	4.340	5.333	41.883	0.181**	-0.067+	0.121**	0.133**	0.161**	0.051	0.122**	0.144**	1

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

Table 4.2 Main Results

Variable	Abandoned Calls		Ambulance Dispatches		Worse Problems	
	(1) OLS	(2) Main Model	(3) OLS	(4) Main Model	(5) OLS	(6) Main Model
Non-Clinical Labor Mix	-0.0012**	-0.0015**	0.0003*	0.0003*	0.0049**	0.0106**
	(0.0003)	(0.0004)	(0.0001)	(0.0001)	(0.0015)	(0.0039)
Offered Calls	0.0000	0.0000	-0.0000**	-0.0000**	0.0000**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Answered Calls	-0.2339**	-0.2346**	0.0664**	0.0663**	0.0570	0.0646
	(0.0302)	(0.0303)	(0.0104)	(0.0104)	(0.0421)	(0.0431)
Transferred Calls	-0.1357**	-0.1307**	0.0024	0.0030	-0.0729	-0.0489
	(0.0321)	(0.0322)	(0.0110)	(0.0111)	(0.0465)	(0.0496)
Transfer Waiting Time	0.0076**	0.0075**	0.0014**	0.0014**	-0.0042*	-0.0043*
	(0.0016)	(0.0016)	(0.0005)	(0.0005)	(0.0021)	(0.0021)
Episode Length	0.0010+	0.0010+	0.0005**	0.0005**	-0.0014+	-0.0014+
	(0.0005)	(0.0005)	(0.0002)	(0.0002)	(0.0008)	(0.0008)
Constant	0.3073**	0.3087**	0.0046	0.0048	0.0371	0.0114
	(0.0313)	(0.0313)	(0.0107)	(0.0107)	(0.0391)	(0.0429)
Observations (N)	642	642	642	642	459	459
R^2	0.469	0.468	0.378	0.378	0.153	0.124
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Center Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

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

their problem got worse. Hence this increase is translated in 9,105 patients experiencing worse problems after using the helpline for every call center per month.

4.6 Robustness Checks

First, we repeat our analysis after substituting our quarter dummies with monthly ones in order to control better for time fixed effect and specific months' fixed effects and find full support for all our hypotheses. Moreover, we test the sensitivity of our results by removing the first three months that every call center has been operating and repeat our analysis. This way we exclude the initial period each call center started operating during which it could experience a number of problems. Our results indicate full support for all our hypotheses.

Next, we check our first hypothesis using a different dependent variable. Specifically, we replace the number of abandoned calls in our model with the ones that were abandoned while talking to a call handler and expect Non-clinical labor mix to decrease them. Similar to our first hypothesis we believe that a high number of call handlers will decrease these abandoned calls. We also divide this variable by the total number of calls offered of each call center every month in order to enable comparison across all call centers. Our results (table 4.3) confirm our expectation since the coefficient of Non-clinical labor mix is significant at 1% and negative providing support for our first hypothesis.

We also test our second hypothesis using a different dependent variable. Specifically, we use the number of patients that were not recommended attending any service but were given self-care advice. If an ambulance is not dispatched for the patient she will be recommended to attend an NHS service or will be given self-care advice and if her situation gets worse then proceed and attend a relevant health service. We believe that call handlers will be less likely to recommend self-care advice due to their limited medical knowledge and risk averseness as we also explained above. As a result, they will be more likely to recommend attending a health service increasing therefore overall cost for NHS. Similar to ambulance dispatches we divide the number of patients that were not recommended attending a health service by the total number of calls offered of each call center every month in order to enable comparison across all call centers. Our results (table 4.4) provide support for our second hypothesis since the coefficient of Non-clinical labor mix is significant at 5% and negative.

Table 4.3 H1 with a different Dependent Variable

Variable	Terminated Calls
Non-Clinical Labor Mix	-0.0013**
	(0.0003)
Offered Calls	-0.0000*
	(0.0000)
Answered Calls	0.0768**
	(0.0224)
Transferred Calls	-0.1338**
	(0.0239)
Transfer Waiting Time	0.0084**
	(0.0012)
Episode Length	-0.0011**
	(0.0004)
Constant	0.0094
	(0.0232)
Observations (N)	642
R^2	0.436
Quarter Fixed Effect	Yes
Center Fixed Effect	Yes

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

Table 4.4 H2 with a different Dependent Variable

Variable	Self-Care Recommendations
Non-Clinical Labor Mix	-0.0008*
	(0.0004)
Offered Calls	-0.0000**
	(0.0000)
Answered Calls	0.1777**
	(0.0297)
Transferred Calls	-0.1781**
	(0.0316)
Transfer Waiting Time	0.0041**
	(0.0015)
Episode Length	0.0010+
	(0.0005)
Constant	0.0283
	(0.0307)
Observations (N)	642
R^2	0.311
Quarter Fixed Effect	Yes
Center Fixed Effect	Yes

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

In addition, we also repeat our analysis for H3 using a different dependent variable. Specifically, we use the percentage of patients that reported that their problem either got worse or remained the same after using the medical helpline and create a new variable called *Unsolved Problems* regarding the health outcome of the call. We expect Non-clinical labor mix to have a positive effect on this new variable decreasing therefore the quality of the provided service. Our results (table 4.5) indicate partial support for our third hypothesis since the coefficient of Non-clinical labor mix is significant at 10% and positive.

Finally, in order to deal with the fact that our dependent variable for our third hypothesis is at a lower frequency than our other variables we interpolate *Worse Problems* from a semi-annual to a monthly frequency (Meijering 2002). In order to do so we use the percentage of calls that resulted in the patient given health information as an indicator. Specifically, such patients were not recommended to contact any other service but were given service location information with the advice to get in touch with them if their situation gets worse. We therefore believe that such patients are highly likely to experience worse health outcomes despite having used the medical helpline which makes them a good indicator for linear interpolation. Hence we create a new monthly variable capturing the percentage of patients who experienced worse health outcome after having used the service using interpolation and expect Non-clinical labor mix to have a positive effect on that. Our results (table 4.6) provide partial support for our third hypothesis since Non-clinical labor mix is significant at 10% and positive. The last two robustness checks for our third hypothesis provide partial support for it. Hence despite our theoretical framework for H3 we do not find full support for it.

Table 4.5 H3 with a different Dependent Variable

Variable	Unsolved Problems
Non-Clinical Labor Mix	0.0093+
	(0.0052)
Offered Calls	0.0000**
	(0.0000)
Answered Calls	0.0266
	(0.0569)
Transferred Calls	-0.0338
	(0.0656)
Transfer Waiting Time	-0.0008
	(0.0028)
Episode Length	-0.0037**
	(0.0011)
Constant	0.2286**
	(0.0567)
Observations (N)	459
R^2	0.095
Quarter Fixed Effect	Yes
Center Fixed Effect	Yes

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

Table 4.6 H3 with Linear Interpolation

Variable	Worse Problems
Non-Clinical Labor Mix	0.0069+
	(0.0042)
Offered Calls	0.0000
	(0.0000)
Answered Calls	0.0027
	(0.0487)
Transferred Calls	-0.0242
	(0.0515)
Transfer Waiting Time	-0.0068**
	(0.0022)
Episode Length	0.0008
	(0.0008)
Constant	0.0883+
	(0.0473)
Observations (N)	357
R^2	0.080
Quarter Fixed Effect	Yes
Center Fixed Effect	Yes

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

4.7 Discussion and Conclusions

Despite their extensive use, little is known about the effect of non-medical workers on healthcare operations. At the same time medical helplines are becoming increasingly popular among such organizations (Spiegelman 2000, Rhian and Claudio 2003) and are expected to offer round-the-clock high quality services (Gustafson 1999). This is translated in low waiting times combined with medical advice and solutions for patients. So in this study we examine the effect of non-medical workers on the service quality of a medical helpline providing useful insights in terms of labor mix strategies. Our dataset allows us to investigate the effect of such individuals on different aspects of the helpline and despite the fact that non-medical workers are considered to decrease overall cost for the organization our results indicate that the relationship is far more complicated.

Specifically, our study reveals a number of interesting results and conclusions. We show that while non-clinical staff members increase the efficiency of patient service by reducing abandoned calls, their broader efficiency benefits are likely to be limited due to other parallel mechanisms which may result in new inefficiencies in other parts of the healthcare system. One such mechanism is the inefficient use of costly resources such as ambulance dispatches. Our results indicate that higher ratio of non-clinical personnel is associated with a higher number of ambulance dispatches: A 50% increase of this ratio results in 15 more monthly ambulance dispatches for each call center. This suggests that call handlers create a significant additional cost for NHS through dispatching more ambulances. While some of these dispatches could be appropriate course of actions in given situations, our results demonstrate a statistically significant relationship between non-clinical labor mix and ambulance dispatches after controlling for endogeneity, indicating potential inefficient use of costly resources with higher non-clinical labor mix. In addition, we also test this hypothesis with a different dependent variable which reveals that call handlers will not recommend self-care to patients increasing therefore overall cost for NHS.

Moreover, while non-clinical workers such as call handlers cost much less in terms of direct labor cost, our results show that they tend to lower quality outcome of the patient service. Especially for National Health Organizations such as NHS this result indicates that call handlers can increase their operating costs in general since a patient whose problem gets worse will definitely use another service of the organization, creating therefore an additional cost for it. Specifically, according to our model a 50% increase in Non-clinical labor mix results in 9,105 patients experiencing worse problems per month despite having used 111 for every call center. These patients will definitely attend another service of NHS increasing therefore overall cost.

As in all empirical studies, our findings and conclusions are subject to limitations. First, our dataset includes no information regarding the patients' condition when they call the medical helpline. Ideally, we would like to have more detailed information about patients' condition, but this was not available in our dataset. In addition, our dataset comes from a single medical helpline. One should be therefore careful when interpreting the results of our study. Despite the fact that medical helplines constitute an appealing setting to study the effect of non-medical workers on the efficiency, cost and quality, generalizing our results to other settings requires a cautious approach.

Despite its limitation, our study extends healthcare operations literature with a new focus on non-clinical workforce in innovative health services. By simultaneously investigating efficiency and quality of patient service, our results highlight potential trade-offs regarding employment of non-clinical workforce. Non-clinical staff members constitute cheaper workforce with compare to medically trained clinical advisors and can therefore be very attractive for health organizations for efficiency gains and direct cost reductions. However, our results suggest a more cautious approach in employing non-clinical workforce as this could potentially create a detrimental effect on the efficiency and quality of healthcare services. Higher non-clinical labor mix may lead to new inefficiencies and associated costs for the healthcare providers through misuse of critical resources (i.e., misuse of ambulance dispatches) or through patients' use of other or additional services if their problem is not resolved satisfactorily in the first instance.

4.8 References

Aksin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665-688.

Aksin, Z., N. Cakan, F. Karaesman, E. L. Ormeci. 2015. Flexibility structure and capacity design with human resource considerations. *Production and Operations Management* **24**(7) 1086-1100.

Anderson, D., G. Gao, B. Golden. 2014. Life is all about timing: An examination of difference in treatment quality for trauma patients based on hospital. *Production and Operations Management* **23**(12) 2178-2190.

Anderson, A., M. Roland. 2015. Potential for advice from doctors to reduce the number of patients referred to emergency departments by NHS 111 call handlers: Observational study. *BMJ Open* **5**(11)

Andritsos, D., C. S., Tang. 2014 Linking process quality and system usage. *Production and Operations Management* **23**(12) 2163-2177.

Appelbaum, P. S., L. H. Roth. 1983. Patients who refuse treatment in medical hospitals. *The Journal of American Medical Association* **250**(10) 1296-1301.

Armony, M., E. L. Plambeck, S. Seshadri. 2007. Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (Why you shouldn't shout at the DMV). *Manufacturing and Service Operations Management* **11**(1) 19-32.

Bang, M., E. L. Ragnemalm. 2012. Persuasive technology: Design for health and safety. *Proceeding of 7th International Conference on Persuasive Technology, Persuasive* 2012, Linkoping, Sweden.

Bhulai, S., G. Koole, A. Pot. 2008. Simple methods for shift scheduling in multiskill call centers. *Manufacturing & Service Operations Management* **10**(3) 411-420.

Bloom, N., J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* **122**(4) 1351-1408.

Bostrom, J, J. Zimmerman. 1993. Restructuring nursing for a competitive health care environment. *Nursing Economics* **11**(1) 35-41.

Briñol, P., R. E. Petty. 2009. Persuasion: Insights from the self-validation hypothesis. *Advances in experimental social psychology* 41 69-118, Zanna, M.P. (eds.), Elsevier, New York, NY.

Brock, D. W., S. A. Wartman. 1990. When competent patients make irrational choices.

New England Journal of Medicine **322**(22) 1595-1599.

Buchan, J., M. R. Dal Poz. 2002. Skill mix in the health care workforce: Reviewing the evidence. *Bulletin of the World Health Organization* **80**(7) 575-580.

Campbell, W. K., A. S. Goodie, J. D. Foster. 2004. Narcissism, Confidence, and Risk Attitude. Journal of Behavioral Decision Making **17**(4) 297-311.

Cayirli, T., E. Veral, H. Rosen. 2008. Assessment of patient classification in appointment system design. *Production and Operations Management* **17**(3) 338-353.

Cayirli, T., K. K. Yang, S. A. Quek. 2012. A universal appointment rule in the presence of no-shows and walks-ins. *Production and Operations Management* **21**(4) 682-697.

Chow, V. S., Puterman, M. L., Salehirad, N., Huang, W., Atkins, D. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management* **20**(3) 418-430.

Clark, J. K., T. Duane, T. Wegener, M. M. Habashi, A. T. Evans. 2012. Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality and Social Psychology Bulletin* **38**(1) 90-100.

Clark, J. K., A. T. Evans. 2014. Source credibility and persuasion: The role of messga position in slef-validation. *Personality and Social Psychology Bulletin* **40**(8) 1024-1036.

Connelly, J. E., C. Campbell. 1987. Patients who refuse treatment in medical offices. *Archives of Internal Medicine* **147**(10) 1829-1833.

Department of Health. 2000. The NHS Plan: A plan for investment, a plan for reform London: Department of Health.

Dubois, C. A., D. Singh. 2009. From staff-mix and beyond: Towards a systemic approach to health workforce management. *Human Resources for Health* **7**(1) 87.

Faulkner, K., A. Dolan, J. White, P. Bentley. 2015. NHS 111 whistleblower speaks: Three weeks' training and I was making life or death decision with no one to turn to for help. *MailOnline.*

Garfink, C, K. K. Kirby, S. S. Bachman. 1991 The University Hospital nurse extender program: Part IV, What have we learned. *Journal of Nursing Administration* **21**(4) 26-31.

Gottlieb, J. B., D. Sarel. 1991. Comparing advertising effectiveness: The role of involvement and source credibility. *Journal of Advertising* **20**(1) 38-45.

Green, L. V., S. Savin, N. Savva. 2013. Nursevendor Problem: Personnel Staffing in the Presence of Endogenous Absenteeism. *Management Science* **59**(10) 2237-2256.

Gustafson, B. M. 1999. A well-staffed PFS center can improve patient satisfaction. *Journal of the Healthcare Financial Management Association* **53**(7) 64-66.

Hovland, C. I., I. L. Janis, H. H. Kelley. 1953. Communication and persuasion: Psychological studies of opinion and change. Yale University Press, New Haven, CT.

Hovland, C. I., W. Weiss. 1951. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly* **15**(4) 635-650.

He, B., F. Dexter, A. Macario, S. Zenios. 2012. The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Operations Management* **14**(1) 99-114.

Hull, L. 2013. New NHS phoneline “will put lives at risk”: Doctors’ warning after 111 number goes into meltdown. *MailOnline.*

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.

Kesavan, S., BR. Staats, W. Gilland. 2015. Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Sci*ence **60**(8) 1884-1906.

Kuo, Y., J. M. Y. Leung, C. A. Yano. 2014. Scheduling of multi-skilled staff across multiple locations. *Production and Operations Management* **23**(4) 626-644.

Krueger, N., J. Norris, P. R. Dickson. 1994. How believing in ourselves increases risk taking: Perceived. *Decision Sciences* **25**(3) 385-400.

Li, L., W. C. Benton. 2006. Hospital technology and nurse staffing management decisions. *Journal of Operations Management* **24**(5) 676-691.

McGinnies, E., C. D. Ward. 1980. Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin* **6**(3) 467-472.

Meijering, E. 2002. A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE* **90**(3) 319–342.

NHS 2010. An overview of NHS Pathways Training. October 2010.

NHS. 2011. Ambulance Services, England 2010-11. NHS Information Centre for Health and Social Care. June 2011.

NHS. 2012. 111 minimum dataset providers version. NHS 111 Programme. Version 0.9. October 2012.

NHS. 2014. NHS 111 Commissioning Standards. NHS England, NHS 111 with CCGs. June 2014.

NHS. 2014. NHS 111 Commissioning Standards. NHS England, NHS 111 with CCGs. June 2014.

Petty, R. E., J. T. Cacioppo, R. Goldman. 1981. Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology* **41**(5) 847-855.

Powers, P, C. Dickey, A. Ford. 1990. Evaluating an RN/co-worker model. *Journal of Nursing Administration* **20**(3) 11-15.

Rhian, S., S. Claudio. 2003. New service design in the NHS: An evaluation of the strategic alignment of NHS Direct. *International Journal of Operations and Production Management* **23**(3/4) 401-417.

Saltzman, R. 2005. A hybrid approach to minimizing the cost of staffing a call center. *International Journal of Operations and Quantitative Management* **11**(1) 1–14.

Saltzman, R. M., V. Mehrotra. 2007. Managing trade-offs in call center agent scheduling: Methodology and case study. *Proceedings of the 2007 Summer Computer Simulation Conference* 643–651, G. A. Wainer, H. Vakilzadian (eds.), San Diego, CA.

Salzarulo, P. A., K. M. Bretthauer, M. J. Cote, K. L. Schultz. 2011. The impact of variability and patient information on health care system performance. *Production and Operations Management* **20**(6) 848-859.

Siebert, W. S., N. Zubanov. 2010. Management economics in a large retail company. *Management Science* **56**(8) 1398-1414.

Smith, C. T., J. De Houwer, B. A. Nosek. 2012. Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin* **39**(2) 193-205.

Spiegelman, P. 2000. Outsourcing your call center. Five benefits to expect if you outsource. *Health Management Technology*.

Staiger, D., J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3) 557-586.

Swindell, J. S., A. L. McGuire, S. D. Halpern. 2010. Beneficent persuasion: Techniques ethical guidelines to improve patients' decision. *Annals of Family Medicine* **8**(3) 260-264.

Tan, T. F, S. Netessine. 2014. When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* **60**(6) 1574-1593.

Telford, L., C. Newell, E. Malnick. 2015. NHS 111 investigation: Handling life-or-death calls at the National Health's non-emergency call centre. *The Telegraph*.

Tormala, Z. L., P. Briñol, R. E. Petty. 2006. When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology* **42**(5) 684-691.

Tucker, A. L, S. J. Spear. 2006. Operational failure and interruptions in hospital nursing. *Health Services Research* **41**(3p1) 643-662.

Turner, J., H. Snooks, A. Youren, S. Dixon, D. Fall, S. Gaze, J. Davies. 2006. The costs and benefits of managing some low priority 999 ambulance calls by NHS Direct nurse advisers. Final report to the Service Delivery and Organisation R&D Programme.

World Health Organization. 2000. Health Systems: Improving Performance. *The World Health Report.*

Ziegler, R., M. Diehl, A. Ruther. 2002. Multiple source characteristics and persuasion: Source inconsistency as a determinant of message scrutiny. *Personality and Social Psychology Bulletin* **28**(4) 496-508.

Chapter five – Conclusions

The increasing importance of knowledge intensive operations and employees in modern economy highlights the need for managers to adopt new team, task and workforce allocation policies that will lead to increased organizational productivity and performance. This thesis contributes to this goal by investigating factors that can impact individual, team and organizational performance. First, it examines the effect of exposure to variety on individual productivity and how it can lead to differentiated effects through different mechanisms. Second, it investigates the effect of past shared experiences of individuals on team productivity by introducing new related concepts and metrics. Finally, it shows how non-clinical workers can affect the performance of healthcare organizations.

It is the author's belief that there are many avenues for further research related to the current thesis. With an increasing amount of data readily available on white-collar operations, empirical studies examining such settings have become popular among Operations Management researchers. This provides many exciting opportunities to better understand and improve such operations, but at the same time it comes with challenges. The introduction and development of analogous constructs of traditional blue-collar concepts to white-collar ones is necessary in understanding the principles under which knowledge intensive organizations operate and in proposing efficient solution approaches and strategies. This thesis constitutes a step towards demonstrating how different task, team and workforce allocation decisions can affect operational performance. In order to achieve this, I introduce mechanisms through which these task, team and workforce dynamics can influence productivity and performance.

Data-driven studies that extend Operations Management literature can ultimately have a significant practical effect on organizations as well. With data becoming more accessible to both companies and academic researchers, and with ever-increasing constraints and pressures on limited organisational resources and intense competition; those organisations which can make use of their data to better understand and improve their operations will have a significant advantage. This provides an excellent opportunity to the scientific community to work on highly relevant, managerially significant, and impactful operational problems which can contribute to both scientific literature and real-life businesses alike.