I, Lauren F. Laker, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Business Administration.

It is entitled:
The Effects of Interruptions and Information Overload on Decision-Making Performance in Knowledge-Work

Student’s name: Lauren F. Laker

This work and its defense approved by:

Committee chair: Craig Froehle, Ph.D.

Committee member: Christopher Lindsell, Ph.D.

Committee member: Jaime Newell, Ph.D.
The Effects of Interruptions and Information Overload on Decision-Making Performance in Knowledge-Work

A dissertation submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Operations, Business Analytics, and Information Systems of the College of Business

By

Lauren F. Laker

Bachelor of Arts, Mathematics and Statistics, Miami University Bachelor of Science, Secondary Mathematics Education, Miami University Masters of Business Administration, Xavier University

October 23, 2015

Committee Chair: Craig M. Froehle, PhD

Committee Members: Christopher J. Lindsell, PhD Jaime B. Windeler, PhD
ABSTRACT

Research suggests that interruptions during a cognitive task can affect the quality and timeliness of decision-making in knowledge-intensive work environments. Moreover, information overload can lead to lower-quality and slower decision-making. This research introduces and tests “emphasis framing” as an operational tactic to help mitigate the effects of information overload and interruptions on the quality and timeliness of decision-making in knowledge-intensive work environments. A series of three experiments was conducted with the following participants: students, crowdsourcing participants, and emergency department physicians. In our studies with students and crowdsourcing participants, while our results were interesting, we were unable to attain statistically significant results. But the results of these experiments did illustrate that studying the complex cognitive tasks associated with knowledge work is nontrivial and highlighted the unique challenges introduced by the knowledge workers themselves and needs to be further explored. Additionally, the experiment with the crowdsourcing participants illustrated some of the challenges with conducting behavioral, knowledge-intensive experiments on crowdsourcing sites and highlighted that further research is needed to determine if that platform is appropriate for these types of experiments. We did attain statistically significant results on our experiment with emergency department physicians. We measured the effect of emphasis framing on two operational performance metrics when under information overload: (1) the quality (accuracy) of the physician’s clinical evaluation, and (2) the efficiency (timeliness) of his/her clinical decision-making. Our results showed that emphasis framing helped mitigate the effects of information overload and increased the quality of
clinical decision-making. Contrary to what we expected, we found that decision-making took longer with the emphasis frame. While we had hypothesized that it would enable the participant to navigate the EHR more quickly, it appears that that perhaps there is actually a quality and timeliness tradeoff such that faster decision-making actually impedes the careful consideration given to high-quality decision making.
ACKNOWLEDGEMENTS

Completing my dissertation and my degree has been an amazing experience that wouldn’t have been possible without the team of mentors, family, and friends who supported me along the way. First and foremost, I would like to thank my husband Jeff. He not only mentally and emotionally supported me on this journey, but was the financial provider for our family, Super Dad to our children, and all around care-taker of our home. From laundry, to cleaning, to cooking, he did it all while I focused on my school work. I could not have done this without your unwavering support and love, I am eternally grateful. I would also like to thank my children, Braden and Abby, for their understanding and support these last few years, your sacrifices and flexibility did not go unnoticed. I love you to the moon and back. And to my in-laws and my dear YaYas, thank you for everything you did for our children. I could begin to count the number of times you helped us in a pinch, went the extra mile, and made all of this possible by feeding and caring for my children while I was working.

And with deepest gratitude, I would like to thank my advisor, Craig Froehle. Thank you for your patience, guidance, support, and inspiration. Thank you for dealing with my stubborn personality and continually pushing me to be better. It was truly an honor to work with you. I would also like to thank my committee members, Christopher Lindsell and Jaime Windeler. Chris, thank you for the amazing opportunity to work with the emergency department the last few years, and even more importantly, for your time and continual support. Jaime, thank you for providing guidance and keen insights. And finally, a special thanks to my entire OBAIS department. Thank you for sharing your time and knowledge so that I could learn from your experience and expertise.

This work is dedicated to Jeff, Braden, and Abby Laker.
# Table of Contents

List of Figures...........................................................................................................................................ix

List of Tables..................................................................................................................................................x

1. Introduction and Motivation.......................................................................................................................1

2. Theory Development and Conceptual Model............................................................................................9
   2.1 Theoretical Foundations & Hypotheses.................................................................................................9
      2.1.1 Knowledge-Intensive Work & the Cognitive Process.................................................................10
      2.1.2 Media Synchronicity Theory........................................................................................................12
      2.1.3 Interruptions.....................................................................................................................................14
      2.1.4 Information Overload......................................................................................................................16
      2.1.5 Emphasis Framing............................................................................................................................17
   2.2 Conceptual Model..................................................................................................................................21

3. Pilot Experiments......................................................................................................................................24
   3.1 Introduction..........................................................................................................................................24
   3.2 Data and Methods................................................................................................................................26
      3.2.1 Experimental Design.......................................................................................................................28
      3.2.2 Student Experiment..........................................................................................................................31
      3.2.3 Crowdsourcing Experiment............................................................................................................33
   3.3 Results and Conclusions.......................................................................................................................37
      3.3.1 Student Experiment..........................................................................................................................37
      3.3.2 Crowdsourcing Experiment............................................................................................................40
   3.4 Discussion.............................................................................................................................................46
4. The challenges of studying the complex and cognitively challenging task associated with knowledge work…………………………………………………………53
   4.1 Introduction………………………………………………………………………….54
   4.2 Background…………………………………………………………………………57
       4.2.1 Knowledge Work…………………………………………………………57
       4.2.2 Experiments with Knowledge Workers………………………………62
   4.3 Discussion…………………………………………………………………………74
   4.4 Appendix………………………………………………………………………….76
   4.5 References…………………………………………………………………………78

5. The operational effect of information overload on clinical decision-making……85
   5.1 Introduction…………………………………………………………………………85
   5.2 Relevant Literature and Hypotheses…………………………………………..87
   5.3 Methods……………………………………………………………………………93
       5.3.1 Sample Design and Selection………………………………………93
       5.3.2 Experimental Protocol and Design…………………………………95
           5.3.2.1 Overview of the Experimental Protocol………………………96
           5.3.2.2 Experimental Interface………………………………………97
           5.3.2.3 Study Design and Development of the Experimental Instrument……………………………………………………………98
           5.3.2.4 Measures……………………………………………………..105
   5.4 Results…………………………………………………………………………….107
       5.4.1 Quality of Clinical Decision-Making…………………………………107
       5.4.2 Timeliness of Clinical Decision-Making……………………………109
### List of Figures

1. Hospitals’ adoption of Electronic Health Record (EHR) systems, 2008-2013……4
2. Information Overload………………………………………………………………………………5
3. Conceptual Model…………………………………………………………………………………23
4. Experimental Design for Pilot Experiments……………………………………………………28
5. Arrangement of Subjects in Student Experiment………………………………………33
6. Student Experiment: Time to complete study, lack of sincere effort…………………38
7. Student Experiment: Time to complete study…………………………………………39
8. Student Experiment: Quality scores……………………………………………………40
9. Mechanical Turk Experiment: Self-reported Effort vs. Attention Check……………43
10.Mechanical Turk Experiment: Time to Complete vs. Self-reported Effort………..44
11.Student Experiment: Data Integrity………………………………………………………….69
12.Mechanical Turk: Attention Check…………………………………………………………..72
List of Tables

1. Pilot Study Scoring Rubric.................................................................31
2. Student Experiment: Study Subjects..................................................38
3. Student Experiment: Time to complete..............................................39
4. Student Experiment: Quality scores.................................................40
5. Mechanic Turk Experiment: Completion and Dropout Rates..............41
6. Mechanical Turk Experiment: Reported Gender..............................41
7. Mechanical Turk Experiment: Reported Ages.................................41
8. Mechanical Turk Experiment: Reported Education...........................42
9. Mechanical Turk Experiment: Accuracy of “Attention Check”............42
10. Mechanical Turk Experiment: Self-Reported Effort..........................43
11. Mechanical Turk: Self-Reported Effort vs. Attention Check...............43
12. Mechanical Turk Experiment: Time to Complete vs. Self-reported Effort.....44
13. Mechanical Turk Experiment: Levene’s Test of Equality of Error Variances......45
14. ED Experiment: Intra-Class Correlation Coefficient........................108
15. ED Experiment: Quality Descriptive Statistics................................108
16. ED Experiment: Levene’s Test of Equality of Error Variances.............108
17. ED Experiment: Timeliness Descriptive Statistics............................109
18. ED Experiment: Levene’s Test of Equality of Error Variances.............109
Introduction and Motivation

Over the previous century, the field of operations management has contributed to a fifty-fold increase in the productivity of the manual worker (Drucker, 1999). But, as we move through the 21st century towards a more knowledge-based economy, there is significant need to further develop the science behind knowledge-work and remove the barriers to efficient and high-quality knowledge-work (Drucker, 1999; Davenport et al., 2002; Ramirez and Nembhard, 2004; Hopp et al., 2009; Froehle & White, 2014). While definitions vary slightly from source to source, knowledge workers have been described as employees with a formal education or high degree of expertise in a particular area, who leverage their knowledge and intellectual capacity to transform information into some form of “product” (Drucker, 1999; Davis, 1999; Davenport et al., 2002; Ramirez and Nembhard, 2004; Hopp et al., 2009). Knowledge work is inherently more cognitive than physical in nature and according to Hopp et al. (2009), “knowledge work is considered a subset of white-collar work, because highly knowledge-intensive tasks are
classified as white collar.” Examples are business and financial operations occupations, such as analysts or accountants, general management occupations, computer and mathematical occupations such as programmers and actuaries, legislators, medical doctors, lawyers, and scientists. There is significant need to shift from the throughput focus that was necessary in the manufacturing-based economy of the 20th-century to focus on developing the science behind knowledge-work and discovering ways to remove the barriers to efficient and high-quality knowledge-work (Drucker, 1999; Davenport et al., 2002; Ramirez and Nembhard, 2004; Hopp et al., 2009; Froehle and White, 2014). This research is focused on further developing the science of decision-making performance in knowledge-intensive work environments.

The healthcare industry is the largest and one of fastest-growing industries in the United States. According to the economists in the Centers for Medicare and Medicaid Services Office of the Actuary, healthcare costs accounted for 17.8% of the gross domestic product of the United States in 2014 and is expected to rise to more than 19.3% by 2023 (Sisko et al., 2014). Healthcare processes are characterized by knowledge-intensive tasks, such as clinical decision-making, and removing the barriers to efficient and high-quality care is of great need. Two landmark reports from the Institute of Medicine (IOM) painted a picture of a healthcare system wrought with quality and efficiency problems and prompted a national focus on healthcare reform. In 1999, “To Err is Human: Building a Safer Health System” exposed the prevalence of medical errors and their implications on the overall cost and quality of healthcare in the United States. In 2001, "Crossing the Quality Chasm” made an urgent call for a redesign of the American healthcare system with a focus on providing high quality (safe, effective,
evidence-based, and patient-centered) care that is delivered in an efficient and equitable manner. A key finding in both reports was the potentially important role of health information technology. A strong healthcare information infrastructure was viewed as critical to facilitating the organized and efficient exchange of information that is necessary to achieve the improvements in quality and efficiency intended by overhauling the U.S. healthcare system. With estimated savings of $81 billion annually from the adoption of interoperable electronic health record (EHR) systems (Hillestad et al., 2005), the federal government committed an unprecedented $27 billion to promote and expand the adoption of health information technology through the enactment of the Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009.

As seen in Figure 1, HITECH has given rise to a greater than 600% increase in the adoption of EHR systems by hospitals since 2008 (Adler-Milstein et al., 2014). EHRs have become the foundation of healthcare information infrastructure in the U.S., but the promises of improvements in information exchange and efficiency have thus far fallen short of expectations (Black et al., 2011; Jones et al., 2012; Kellermann & Jones, 2013). Jones et al. (2012) contend that, “swapping out of the medical record cabinet and prescription pad for a computer is proving insufficient to realize the benefits of health IT” (p. 2244). This apparent contradiction between investment in these operational technologies and lack of realized improvement not only highlights the fact that we do not adequately understand the factors that contribute to efficient and effective work systems, but most alarmingly, the quality of patient care may also be suffering.
Effective operations management requires high-quality and timely decision-making. In knowledge-work environments, such as hospitals and other professional services, interruptions and information overload are two phenomena that increasingly threaten the quality and timeliness of decision making. We introduce and test “emphasis framing” as an operational tactic to help mitigate the effects of information overload and interruptions on the quality and efficiency of decision-making. Emphasis framing occurs when some aspect or component of the information being exchanged is highlighted or stressed to make it more easily, or likely to be, processed by the recipient (Entman, 1993; Druckman, 2001). From a cognitive perspective, a “frame” serves as a simplifying structure for cognitive categorization (Davies & Mabin, 2001), which decreases cognitive load and potentially improves decision-making performance. This research evaluates the effects of emphasis framing as an operational tactic to improve cognitive categorization, decrease cognitive load, and improve the quality and efficiency (timeliness) of decision-making. While this research may be applicable to any
knowledge-intensive decision-making process, it will specifically test the hypothesized relationships in a healthcare work environment.

Information overload occurs when the information intensity increases to a point where the information-processing requirements of a task exceed the information-processing capacities of the individual (Eppler & Mengis, 2004). It is largely determined by the quantity and complexity of information needing to be processed and plays an important role in the ability of a decision-maker to accurately and efficiently process information (Hiltz & Turoff, 1985; Keller & Staelin, 1987; Schneider, 1987; Schick et al., 1990; Speier et al., 1999; Eppler & Mengis, 2004). While the most-informed decision is often the best decision, the theory of bounded rationality argues that humans have only a limited capacity to process complex problems and information (Simon, 1957). As seen in Figure 2, up to a certain point, decision-making performance is positively correlated with the amount of information a decision-maker receives. But, beyond that point, the information-processing requirements of a task exceed the information-processing capacities of the decision-maker, sending him/her into a state of information overload (Eppler & Mengis, 2004). As a result, decision-making performance decreases.

**Figure 2. Information Overload**

![Figure 2. Information Overload](image)
In healthcare, EHRs provide information that could potentially increase the quality and efficiency of clinical decision-making and improve patient care. But, the evidence suggests that EHRs are currently failing to meet those objectives and, even worse, potentially contributing to errors in clinical decision-making (Ash et al., 2004; Kuperman, 2011; Singh et al., 2013). The role of information overload on the timeliness and quality of clinical decision-making performance needs to be better understood.

Prior research does not offer a clear conclusion as to the relationship between information overload and information technology. Some research has touted information technology as a potential countermeasure against information overload (Huber, 1984; Hiltz & Turoff, 1985; Schick et al., 1990; Edmunds & Morris, 2000). For example, Edmunds and Morris (2000) suggest that “push technology”, which pushes information to users based on pre-selected triggers, allows users to reap the benefit of the increased information available while decreasing the amount of information they would consume as compared to a traditional information pull. On the other hand, other research has found that information technology contributes to information overload (Speier et al., 1999; Ash et al., 2004; Eppler & Mengis, 2004; Harrison et al., 2007). Speier et al. (1999) contend that “…information technology may be a primary reason for information overload due to its ability to produce more information more quickly and disseminate this information to a wider audience than ever before” (p. 337). While the research is limited at this time, the conclusion thus far is that health care information technology is contributing to the intensity of information being presented, potentially inducing information overload, rather than alleviating it (Ash et al., 2004; Harrison et al., 2007; Kuperman, 2011; Singh et al., 2013). A 2013 study published in JAMA found that
over one-third of physicians reported missing test results in an EHR system because they are simply overwhelmed by data and information (Singh et al., 2013). EHRs have flooded physicians with data and information that could potentially increase the quality and efficiency of clinical decision-making, but is currently failing to meet those objectives, and could possibly even be contributing to errors in clinical decision-making. Information overload is a problem in healthcare and other knowledge-intensive work environments, a need exists for research to develop a better understanding of how operational policies and system design of information technology can help mitigate the effects of information overload.

Interruptions and decision-maker characteristics, most notably experience, have also been shown to have a direct effect on decision-making performance, and possibly mediates the relationship between the information characteristics and decision-making performance (Speier et al., 1999; Eppler & Mengis, 2004; Tucker and Spear, 2006; Froehle & White, 2014). According to Zellmer-Bruhn (2003), interruptions are described as incidents or occurrences that impede regular work flow. Interruptions affect the way information is processed by the decision-maker and can lead to a decrease in the quality and efficiency of decision-making performance (Wood, 1986; Speier et al., 1999; Eppler & Mengis, 2004; Eppler, 2006; Tucker and Spear, 2006; Kahneman, 2011; Froehle & White, 2014). Many studies have shown that interruptions are frequent and pervasive in clinical practice (Chisolm et al., 2001; Coiera et al., 2002; Tucker and Spear, 2006; Laxmisan et al., 2007; Westbrook et al., 2010). In one study by Chisolm et al. (2001), they reported that physicians are interrupted on average 5.8 times per hour, with emergency physicians on the higher end of the spectrum with an average of 9.7
Interruptions per hour. Interruptions are of particular concern in health care due to their potential influence on clinical decision-making, but prior research has demonstrated they are equally as problematic for decision-making in other knowledge-intensive work environments (Speier et al., 1999; Speier et al., 2003). Interruptions have long been a reality in health care and other knowledge-intensive work environments, and the introduction of information technology only further complicates the situation.

Following this introduction, Chapter 2 includes a discussion of the theoretical foundations of the research and their contribution to the conceptual model and formal hypotheses. Chapter 3 details of the two pilot studies conducted with students and Mechanical Turk. Chapter 4 is written as a stand-alone paper that includes some details of our pilot studies but is written to highlight the emergence of research dedicated to knowledge-work, discusses some of the challenges with studying the complex cognitive tasks synonymous with knowledge work, and provides some general recommendations for those considering laboratory experiments with knowledge workers. Due to the fact that it is drawn from my dissertation, please note there will be some overlap with the theory and development. Chapter 5 details the study conducted with the Emergency Department Physicians. Finally, the last chapter contains a synthesis of the results from all experiments and a discussion of the contributions of our work.
2

Theory Development and Conceptual Model

2.1 Theoretical Foundations

This research contributes to the development of our understanding of decision-making in knowledge-intensive work environments. It builds on the operations management literature and the cognitive sciences to explore the decision-making process in knowledge-intensive work environments. It also incorporates healthcare research on clinical decision-making. It draws from the operations, healthcare, information systems, and cognitive science literature to develop our understanding around the role of interruptions, information overload, and decision-making. Finally, the proposed operational tactic of emphasis framing builds on the concept of framing from the cognitive sciences (Wicks, 1992; Entman, 1993; Davies and Mabin, 2001; Druckman, 2001) and Media Synchronicity Theory (Dennis et al., 2008) to explain the role of information systems in decision-making performance.
2.1.1 Knowledge-Intensive Work & the Cognitive Process

The fields of operations research and management science contributed to a greater than 50-fold increase in the productivity of manual workers over the last century (Drucker, 1999). However, as the economy shifts from depending on the productivity of the manual labor workforce to depending on the productivity of knowledge workers, there is significant need to further develop the science behind knowledge-work and remove the barriers to efficient and high-quality knowledge-work (Drucker, 1999; Davenport et al., 2002; Ramirez and Nembhard, 2004; Hopp et al., 2009; Froehle and White, 2014). But, it is challenging to measure the productivity of the complex cognitive tasks synonymous with knowledge work.

In order to begin developing the science behind knowledge work, Ray and Sahu (1989) helped by defining the types of cognitive tasks often associated with knowledge work. They classify knowledge-intensive tasks into two categories based on their overall complexity and required mental effort. In the first category are routine or repetitive tasks requiring minimal mental effort, such as clinical procedural tasks. The inherent nature of these tasks places limited cognitive demand on the worker but they are still classified as knowledge-intensive due to a specific level of expertise or education that are required for their completion. In the second category are non-repetitive or non-routine tasks that place greater cognitive demand on the worker due to the fact that the tasks generally involve the processing and synthesis of complex, interrelated information. A defining attribute of knowledge workers is their aptitude for the non-repetitive, non-routine cognitive tasks and their ability to generate knowledge and make decisions.
From the healthcare perspective, clinical work is often considered knowledge-intensive but, similar to the operations literature, not all clinical tasks are created equal. According to Li et al. (2012), clinical tasks are generally categorized into three groups: procedural, decision-making, and problem-solving. Procedural tasks are often routine and repetitive and involve the automatic activation of knowledge obtained through procedural training. The problem-solving and decision-making tasks are similar to the non-repetitive, non-routine activities discussed in the operations literature. They involve the deliberate and systematic processing of a variety of clinical evidence to support effective clinical reasoning and medical decision-making.

Daniel Kahneman’s (2011) work on the cognitive process and decision-making has a different theoretical basis than the research in the operations literature, but similar in that he postulates a dual process theory in which the brain uses two fundamentally different systems to process information and make decisions. He posits that activation of System 1 occurs when workers are presented with routine and repetitive tasks and decisions are made quickly and intuitively, such as the act of typing on a computer without thinking about each individual depression of the keys. When an individual is presented with non-routine or non-repetitive tasks (i.e. information-intensive tasks), System 2 is activated and decisions are slower, more deliberate, and often require intense focusing. Additionally, the tasks that require the activation of System 2 require significant cognitive exertion and the decision to activate System 2 must be a deliberate action taken up by the individual. The focus of this discussion is focused on System 2 decision-making that is a necessary for many knowledge-intensive tasks.
To evaluate decision-making performance, we refer to Ramirez and Nembhard’s (2004) work on measuring knowledge-worker productivity. Appropriate and accurate measurement of knowledge-worker productivity can be beneficial for monitoring knowledge workers, assisting with capacity and strategic planning, reducing subjectivity of knowledge-worker evaluations, and helping to establish benchmarks. Different measures capture different dimensions of knowledge-worker productivity and the focus of this research is to focus on quality and timeliness. Quality is central to most aspects of operations management and was selected for this study because it is also one of the most important measures of productivity among knowledge workers. Additionally, in an era when Americans take in five times more information every day than they did 30 years ago (Levitin, 2014), the ability to process the deluge of information in a timely manner is becoming an increasingly important measure of productivity among knowledge workers and was selected as our second focus for this study. Additionally, from a healthcare perspective, both of these measures were highlighted in the IOM’s landmark report “Crossing the Quality Chasm”. In order to improve the quality of care provided, clinicians need to process and synthesize a large variety of clinical information to make high-quality medical decisions, and must do so in a timely manner.

2.1.2 Media Synchronicity Theory

According to research from IBM, “every day we create 2.5 quintillion bytes of data – so much that 90% of the data in the world has been created in the last two years alone.” In today’s information age, knowledge workers regularly interact with a wide variety of information systems. Media Synchronicity Theory (MST) (Dennis et al., 2008)
is helpful in understanding the role of information systems in decision-making performance. MST posits that matching the capabilities of a medium, that best supports the conveyance or convergence process, can affect the cognitive load of a decision maker and, ultimately, the performance of the process.

For the portions of a decision process where conveyance of large amounts of raw information is required, “individuals will have less of a need to transmit and process information at the same time” (Robert & Dennis, 2005) and media that supports low synchronicity is most appropriate. The logic behind this media choice lies in the fact that the conveyance process allows for individual consumption and processing of new or diverse information with decreased interaction between individuals. Additionally, if the conveyance of information is truly needed (vs. convergence) a more synchronous medium can actually increase the cognitive load on a decision maker and negatively influence decision performance. Currently, the design of EHRs is primarily geared toward conveying information about patients and their medical history. Alternatively, more synchronous media, which allows individuals a high level of interaction and the ability to develop a shared understanding of raw information, is most appropriate for convergence as it allows for faster transmission and exchange of information. Convergence generally requires fewer cognitive resources than conveyance and ultimately reduces cognitive effort (Dennis et al., 2008).

In accordance with McGrath’s (1991) time, interaction, and performance (TIP) theory, MST also posits that information requirements change over time. As a result of training and prior experiences, as individuals become more familiar with tasks, social norms, the media, and individuals involved in the process, the exchange of information
will involve more convergence and less conveyance. This theory might indicate that as the use of EHR systems evolves, higher media synchronicity may be more appropriate.

### 2.1.3 Interruptions

In addition to needing to understand the role of information systems in decision-making performance, in order to explain the operational implications of interruptions on the timeliness and quality of decision-making, we draw our knowledge from the operations management and information systems literature. First, interruptions decrease the timeliness of the decision-making process by simply increasing the total length of the decision-making process itself. Second, interruptions increase the cognitive load of the decision-maker (Baron, 1986; Speier et al., 1999; Tucker and Spear, 2006). Due to the limited cognitive-processing capacities of humans (Simon, 1957), the increased cognitive demands are associated with decreased decision-making performance (Speier et al., 1999; Eppler & Mengis, 2004; Tucker & Spear, 2006). Third, the characteristics of the interruption, such as the quantity and complexity of information being communicated, and the similarity of the interruption to the task being interrupted, have all been shown to influence the way in which information is processed (Speier et al., 1999; Eppler & Mengis, 2004). As the duration, quantity, and complexity of information conveyed in an interruption goes up, it has a negative influence on the quality and efficiency of the decision-making process (Speier et al., 1999; Eppler & Mengis, 2004; Tucker & Spear, 2006; Froehle & White, 2014). Upon resumption of an interrupted task, further time and effort for re-work may be necessary and quality may deteriorate as steps or information may inadvertently be missed or
skipped (Wood, 1986; Bailey, 1989; Eppler & Mengis, 2004; Eppler, 2006; Tucker & Spear, 2006; Froehle & White, 2014). Additionally, recall accuracy of information processed prior to an interruption may deteriorate (Baron, 1986; Tucker & Spear, 2006). When the content of the primary task is significantly different than the content of the interruption, the decision-making process takes longer and the quality is negatively influenced as the amount of information needing to be processed often exceeds the information-processing capacity of the decision-maker (Iselin, 1988; Speier et al., 1999).

The richness of the information being transmitted and the objective of the information exchange, conveyance or convergence, also influence information processing. According to Daft and Lengel (1983), information richness, an attribute of the medium, is defined by the amount of detail contained in an exchange of information that helps provide understanding. For example, a text exchanged between friends is considered a lean information exchange in comparison to an information-rich face-to-face exchange due to the fact that, in addition to simple factual information provided in a text, a face-to-face interaction can also convey body language, facial expressions, and intonation. According to Media Synchronicity Theory (Dennis et al., 2008), for a communication that requires conveyance of information, such as an interruption, the use of an information-rich or highly synchronous medium will be more disruptive than a less synchronous medium. Alternatively, if convergence is necessary, an information rich or highly synchronous medium will be more appropriate than a less synchronous medium, which would increase cognitive load and slow down decision-making.

Many factors contribute to the way information is processed by a decision-maker and how it influences the quality and efficiency of decision-making performance. This
research builds on the prior research that suggests interruptions negatively influence decision-making performance in knowledge-intensive work environments, thereby hypothesizing:

*H1:* In knowledge-intensive work environments, interruptions increase decision-making time.

*H2:* In knowledge-intensive work environments, interruptions decrease decision-making quality.

2.1.4 Information Overload

As an ever-increasing quantity of information is being made available by information technology to decision-makers, the individual limits of information-processing capacity are subject to being challenged by the intensity of information, increasing the risk and influence of information overload on decision-making performance. Up to a certain point, decision-making performance is positively correlated with the amount of information a decision-maker receives. But beyond that point, the information overload negatively influences the processing capacity of the decision-maker and decision-making performance decreases (Eppler & Mengis, 2004); it becomes difficult for a decision-maker to identify relevant information and effectively map relationships between key details (Schneider, 1987; Speier et al., 1999; Eppler & Mengis, 2004). Additionally, the efficiency of the decision-making process is hampered because individuals need more time to process the quantity and complexity of information needed to reach a decision (Jacoby, 1984; Hiltz & Turoff, 1985).
To develop an understanding of the factors that induce information overload and affect decision-making performance, we refer to the information systems literature. First, the characteristics of the task have been shown to influence information overload. Eppler and Mengis (2004) found that a high volume of information, task complexity, and novelty of information increase the information-processing requirements of a decision-making task. Second, the characteristics of the decision maker has been shown to influence information overload (Speier et al., 1999; Eppler and Mengis, 2004). The more qualified and experienced a decision-maker, the faster and more efficient he/she is at processing information. Third, interruptions have been shown to have a direct effect on decision-making performance (Speier et al., 1999; Eppler & Mengis, 2004; Tucker and Spear, 2006; Froehle and White, 2014). According to Zellmer-Bruhn (2003), interruptions are described as incidents or occurrences that impede regular work flow. Interruptions affect the way information is processed by the decision-maker and can lead to a decrease in the quality and efficiency of decision-making performance (Wood, 1986; Speier et al., 1999; Eppler & Mengis, 2004; Eppler, 2006; Tucker and Spear, 2006; Kahneman, 2011; Froehle & White, 2014). Interruptions have long been a reality in knowledge-intensive work environments, and the introduction of information technology has added another complicating factor.

2.1.5 Emphasis Framing

There is a significant body of literature on the concept of framing, predominately from the social and cognitive sciences exploring how frames influence political opinion formation. Framing can be categorized into two groups, equivalency framing and
emphasis framing. Equivalency framing looks at how the use of logically equivalent information, which is usually presented in a positive or negative way, influences the preferences of an individual. A classic, widely cited example of equivalency framing comes from Tversky and Kahneman (1981). First, they presented a group of participants with a problem and two quantitatively equivalent solutions. The two equivalent solutions were described in terms of the number of people that could be saved from a disease. Option A was presented as a risk-averse choice with a certain outcome (certain number of lives saved), while Option B was presented as a risk-seeking choice with an uncertain outcome. When being asked to make a choice about saving a person, the participants chose the more certain, risk-averse option 72% of the time. In the second phase of their experiment, they presented the same information to a different group of participants, with one slight change. Instead of presenting the solutions in a “lives saved” frame, they presented two quantitatively equivalent solutions in terms of the number of people that would die from a disease. Option A was again presented as a risk-averse choice with a certain outcome, while Option B was presented as a risk-seeking choice with an uncertain outcome. When participants were asked to make a choice framed in terms of the number of people who would die, the participants chose the risk-seeking option 78% of the time over the certain option. This study showed how, despite presentation of equivalent information, the positive or negative portrayal of information can influence preferences. Research suggests that this type of framing works by priming the subconscious mental process (Levin et al., 1998). In terms of Kahneman’s dual process of decision-making, we would infer that this type of framing effects System 1 decision-making. The participants were biased by the
presentation of the information and exhibited errors in intuitive thought as a result of System 1 thinking. In theory, had the participants consciously enacted their System 2 thinking, they would have inferred that they were presented with equivalent information.

Druckman (2001) expanded on the concept of framing introduced by Tversky and Kahneman (1981), by introducing the concept of emphasis framing. Unlike with equivalence framing, individual preferences are not changed by emphasis framing. Instead, emphasis is made on some aspects of the information that is being exchanged to make it more salient, or more prominent and meaningful, in an attempt to make it more easily processed by individuals (Entman, 1993). An emphasis frame exerts influence over human consciousness by highlighting some information to make it more noticeable, meaningful, or memorable. This increase in salience enhances the probability that an individual will identify and discern meaning from information being communicated and thus process it more readily (Fiske and Taylor, 1991). An example of an emphasis frame would be a research abstract. It provides a summary of the research to the reader and helps to guide their understanding of the problem being studied, the methods employed, study results, and conclusions. From a cognitive perspective, a frame proves beneficial as a simplifying structure for cognitive categorization (Wicks, 1992; Davies and Mabin, 2001), which decreases cognitive load and potentially improves decision-making performance. In light of the rapid expansion of information technology in the healthcare sector and the possibility of information overload, this research evaluates the effect of emphasis framing as an operational policy to improve cognitive categorization, decrease cognitive load, and improve the quality and efficiency of decision-making.
Clinical decision making is a complex, knowledge-intensive process. After the 1999 landmark publication of “To Err is Human: Building a Safer Health System,” the prevalence of medical errors also became much more well-known. According to Croskerry et al. (2013), it is the “flaws in clinical reasoning rather than lack of knowledge that underlie cognitive errors”. Drawing from Kahneman’s dual-process theory of decision making, they conclude that the majority of clinical errors occur “on the front line” when clinicians are under immense time pressures and resources are in short supply. Under those trying conditions, they posit that physicians are subconsciously looking for shortcuts or ways to reduce the excessive cognitive loading and revert to trained, procedural knowledge executed with System 1 thinking, instead of the System 2 mode of thinking that is more appropriate for the complex, knowledge-intensive nature of clinical decision making. For example, a physician’s perception of risk to a patient may be influenced by whether the outcome for a patient is expressed in terms of the possibility that the patient might die or live. Instead of being persuaded by an equivalency frame, physicians should carefully consider all potential outcomes and contingencies of a clinical problem.

In an effort to reduce clinical errors and improve decision-making, Graber et al. (2012) suggest three primary ways physicians can override the intuitive System 1 thinking and move into System 2 thinking: 1) increase their knowledge through disease-specific training and deliberately seeking feedback on past performance; 2) improve clinical reasoning with reflective practice and metacognition; and 3) working with other people and with the assistance of decision-support tools to augment an individual physician’s own knowledge and capabilities. All three of these interventions emphasize
the importance of feedback and convergence on the large volume of complex, interrelated information they are processing in order to produce high quality (safe, effective, evidence-based, and patient-centered) medical decision-making. This research proposes emphasis framing as an operational tool to improve cognitive categorization, decrease cognitive load, and improve the quality and efficiency of decision making is not only aligned with the cognitive sciences but closely aligned with suggestions for improving clinical decision-making in the healthcare literature. Therefore, this research hypothesizes:

\[ H3: \text{In knowledge-intensive work environments, emphasis framing reduces decision-making time.} \]

\[ H4: \text{In knowledge-intensive work environments, emphasis framing improves decision-making quality.} \]

2.2 Conceptual Model

Our conceptual model and hypotheses were developed based on our theoretical foundations that detailed how information overload and interruptions influence decision-making performance. In high-information-intensity environments, the ability of a decision-maker to process information can be overwhelmed. In these situations, a phenomenon called “information overload” occurs. Information overload exists when the information-processing requirements of the decision-making task exceed the information-processing capacity of the decision-maker. We introduce and test “emphasis framing” as an operational tactic to help mitigate the effects of information overload. From a cognitive perspective, an emphasis frame serves a simplifying
structure for cognitive categorization, which decreases cognitive load and potentially improves the quality of decision-making. We posit that emphasis framing will have a positive effect on the quality and efficiency by improving the accuracy and timeliness of decision-making.

The effects of interruptions have been extensively studied, especially in repetitive-task industrial settings. There is less insight available regarding how interruptions affect knowledge-work. Our research incorporates distraction interruptions as a manipulated condition in order to assess their effects on decision-making timeliness and quality. For example, if a physician receives a phone call in the middle of reviewing a patient’s chart, it can both delay completion of the chart review (even more than the duration of the interruption) and introduce decision errors. Previous work using discrete-event simulation has been instrumental in identifying the potential magnitude of interruption-induced forgetting’s effect on knowledge-work efficiency. However, there is still much more to be learned regarding how, and how much, interruptions delay, and introduce errors into, the decision-making process. We posit that interruptions will have a negative effect on the quality and efficiency of decision-making.

In addition to the evaluation of our the effect of our two independent variables on our dependent variables, we will be collecting data to evaluate if there is any moderating effects of two background variables, decision-maker qualifications and gender.
Figure 3. Conceptual Model

Independent Variables

- Emphasis Framing
  - H1, H2 (+)
  - H3, H4 (-)

- Interruptions

High-Information-Intensity Environment

Dependent Variables

Decision-Making Performance:
- Quality (accuracy)
- Efficiency (timeliness)

Background Variables
- Decision-Maker Qualifications
- Decision-Maker Gender
3

PILOT EXPERIMENTS

3.1 Introduction

We designed and executed two pilot experiments to investigate the effects of interruptions and emphasis framing on the quality (accuracy) and efficiency (timeliness) of knowledge-intensive, decision-making where information overload is present. We chose to execute experiments because they are helpful for understanding the many behavioral and operational aspects of how people interact with work systems. They are a noteworthy methodology that has been used for years in the psychology, economics, and behavioral decision research literature, but it is among one of the least-used among the large body of operations research (Bendoly et al., 2006; Fisher, 2007; Boyer and Swink, 2008). While many different methods may be appropriate for investigating knowledge worker productivity, the most appropriate research strategy is always a dilemmatic choice. According to McGrath (1981), research should aim to maximize: generalizability over populations; control of variables; and realism of observational context. But, due to conflicting objectives it must be understood that maximization of
one of these will be done at the expense of the other two. For example, experiments, which systematically derive and analyze data from direct or indirect observations (Roth, 2007), are particularly well suited to investigating knowledge worker productivity given their strength for precision, control, and ability to replicate. Though they offer less external validity than other empirical methods, a well-designed experiment can help parse out the complex tasks associated with knowledge work while abstracting away unnecessary details without a loss of generalization. Empirical research methods, such as experiments, that systematically derive and analyze data from direct or indirect observations (Roth, 2007), are especially useful to test theories about knowledge-worker productivity. Knowledge workers leverage their knowledge and intellectual capacity to transform information into some form of “product.” The generation of this “product” relies on processing of information and ultimately the decision-making of the knowledge workers. Prior research in psychology and economics has shown that humans are inherently irrational and human behavior and cognition affects decision-making (Tversky and Kahneman, 1974; Simon, 1979). In other words, human are influenced by their beliefs, biases, motivations, and cognitive limits and often deviate from logically optimal solutions even when presented with perfect (full and accurate) information. These findings are in distinct contrast to traditional operations management models which often assume rational behavior of decision makers, such that all decisions will serve to maximize some sort of utility function. This contradiction highlights the importance and need for incorporating behavioral factors into operations management and use of empirical methods to help develop the science behind knowledge worker productivity.
3.2 Data and Methods

In our pilot studies, we designed and executed a series of experiments to better clarify the relationships of interest, to gather preliminary data in order to better estimate effect sizes, and to test the experimental technology (video- and browser-based simulation of a complex decision-making task).

The first experiment was a controlled laboratory experiment conducted with undergraduate business students recruited from the University of Cincinnati. While students are often a convenience sample in laboratory experiments, they are also a logical choice for a study focused on the highly cognitive tasks synonymous with knowledge work. They regularly participate in knowledge-intensive tasks; students are taking in new information, synthesizing it, and generating new knowledge when they complete their coursework, sit for exams, and even when they participate in classroom discussions. Additionally, while the use of students has often been a debated aspect of laboratory experiments, several studies have shown no statistically significant difference between the performance of students and members of the representative population under study, usually managers (Plott, 1987; Ball and Cech 1996; Bolton et al. 2012; Moritz et al., 2013).

The second experiment was conducted online with subjects recruited from Amazon’s crowdsourcing site, Mechanical Turk (MTurk). Crowdsourcing sites, such as Mechanical Turk, were initially developed for use by companies primarily seeking to outsource remedial tasks that computers are currently unable to do, such as transcribe books/videos or classify images. In general, this is how MTurk works: A “Requestor”
(employer) posts a job called a “HIT” (Human Intelligence Task) and “Workers” (employee) can then choose to participate in any HIT for which they are qualified to participate\(^1\). These sites have quickly drawn the appeal of researchers for their potential for providing quick and easy access to online research participants, minimal costs (recruitment and administrative), and a more diverse participant population than normally seen in typical student samples. But, there has been some concern among researchers with potential quality issues related to “fast and cheap” data. While only a very minimal number of studies addressing the quality concerns thus far, much of the research suggests that despite the concerns, there is potential for Mechanical Turk to obtain high-quality data inexpensively and rapidly (Paolacci et al., 2010; Buhrmester et al., 2011; Mason & Siddharth, 2012; Rand, 2012; Chandler et al., 2014) as long as researchers understand its limitations and managing potential risks. The following are some of their recommendations for ensuring high-quality data results from behavioral research on Mechanical Turk: 1) track subjects via their unique worker ID to ensure independent responses (Paolacci et al., 2010); 2) recognize that participants share information and interact in online communities such as mturkforum.com, Reddit, and Facebook (Chandler et al., 2014); 3) consider employing some form of Frederick’s Cognitive Reflection Test (2005) to gauge cognitive effort (Goodman et al., 2013); and 4) include some form of an “attention check”\(^2\) to assess whether participants carefully

---

\(^1\) Amazon Mechanical Turk allows Requestors to qualify users before they work on their HITs. The qualifications can be anything from a gender or location-based requirement to responding to a series of questions or have a specific profile on Mechanical Turk based on their historical performance on HITs.

\(^2\) Two common “attention checks” used in these studies: 1) the Instructional Manipulation Check (Oppenheimer et al., 2009) which asks a question at the end of the study (such as “What was this study about?”) to gauge whether participants carefully read the instructions and 2) Reverse Turing test questions somewhere in the midst of the study (such as “Answer yes, if you are reading this question.”).
read instructions (Mason & Suri, 2012; Goodman et al., 2013; Chandler et al., 2014). As Amazon’s Mechanical Turk is a relatively new source of potential participants for experiments, it was used for this study as a means to test whether our hypotheses hold in another population.

3.2.1 Experimental Design

Both experiments were designed as balanced, fixed-effects means models, with a complete 2-way factorial treatment structure (Figure 4). The survey was deployed using Qualtrics survey software (Provo, Utah: v. 2009).

**Figure 4. 2x2 Experimental Design**

Group A was a control group that was presented with all information needed to make the decision, but did not receive an interruption or emphasis frame. Group B was presented with the same information as Group A, but respondents were interrupted by a video that was intended to convey a large quantity of complex information dissimilar to the primary knowledge-intensive decision-making task. Group C received the same information as Group A, except respondents were presented with a video that provided emphasis framing. The information in the “emphasis frame” was not new information, the same information was provided in the written information provided to all participants,
it was simply emphasized in the video presented to the participants in Groups C and D. Group D received the same information as Group A, along with the interruption from Group B and the emphasis frame from Group C.

Respondents were presented an “on-line dating” scenario with information being presented over multiple pages within the Qualtrics survey tool. Subjects analyzed and processed a large volume of information to determine the day, meal, and location for two individuals (Adeline and Declan) to meet that was optimally aligned with their preferences and minimized their total travel expenses. First, the subjects were presented with a description of the problem (Appendix A1). Second, the subjects were presented with basic dating profiles for Adeline and Declan (Appendix A2). A large amount of information was presented in order to attempt to induce information overload in the subjects. Key information in Adeline and Declan’s profiles about their dining preferences needed to be parsed out to enable them to make the optimal decision at the end of the experiment. For Groups C and D, the subjects were presented with a video of Adeline who placed an emphasis on her preference for Mexican food. Adeline was portrayed in the video by a paid actress.

Third, the subjects were presented with calendars for Adeline and Declan (Appendix A3). The only day/time/meal combination available on both schedules is Saturday evening. The fourth set of information presented to the subjects was the description of the 10 cities, including restaurant choices, in the fictional country of Zen where Adeline and Declan reside (Appendix A4). Four of the cities do not have restaurants, leaving 6 cities to choose from for their first date. After the city descriptions, Groups B and D were presented with an interruption. It consisted of a video detailing the
classic “Monty Hall” probability puzzle. The interruption engaged subjects in a cognitive task that was knowledge-intensive but dissimilar in content to the primary task. Finally, all groups were presented with a train schedule (Appendix A5). The information enabled them to determine that traveling to Delta, Gamma, or Upsilon would minimize travel costs. Since there are no restaurants in Delta, the only two feasible options were Gamma and Upsilon. When considering the preferences of Adeline and Declan, in combination with the restaurant information in the city descriptions, the El Habanero restaurant in Gamma was the optimal choice.

In order to evaluate the timeliness of the decision-making, the survey tool recorded information on how long a subject spent on each page. In order to evaluate the quality of the decision-making, a scoring rubric was developed with a minimum score of zero and maximum score of nine (See table below). The three key variables in this experiment were as follows: day of week and time of day for a meal; city; and restaurant. First, when evaluating their schedules, the only day of the week that works for both Adeline and Declan is Saturday. While lunch is potentially a feasible choice for meeting, it is not optimal as Adeline has a prior obligation that keeps her busy until noon at an undisclosed location. Dinner is the optimal time for both of them to meet as neither has plans Saturday late afternoon or evening. Therefore, 1 point was awarded for selecting to meet for lunch on Saturday and 3 points were awarded for selecting to meet for dinner on Saturday. Next, the city and train schedule information needed to be considered concurrently. When evaluating the train schedules, there were only two feasible cities that minimize cost and have dining options: Gamma and Upsilon. When considering the restaurants available in these cities, there were three feasible choices in
Gamma and two feasible choices in Upsilon. But based on the collection of information provided, El Habanero in Gamma was the optimal choice with Marni Thai in Upsilon being the second best choice. Therefore, 6 points were awarded for the optimal city/restaurant combination, 4 points for an optimal city and feasible but sub-optimal restaurant combination, 2 points for the two feasible but sub-optimal city/restaurant choices, and 0 points for all other combinations. The scoring rubric for the pilot experiments is presented in Table 1.

**Table 1. Pilot Study Scoring Rubric**

<table>
<thead>
<tr>
<th>Category</th>
<th>Combination</th>
<th>Pts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day/Meal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturday/Dinner</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Saturday/Lunch</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td><strong>City/Restaurant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma/El Habanero</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Upsilon/Marni Thai</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Gamma/Fog Harbor</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Gamma/Grimaldi’s</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Upsilon/Bin 36</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

### 3.2.2 Student Experiment

The data collection for the student experiment took place over a 4-week period and was conducted solely by the principal investigator. The target subjects (n=28) were undergraduate students accepted into the Lindner College of Business at the University of Cincinnati. Given the admission requirements for the Lindner College of Business, we could reasonably assume that these subjects would possess the skills necessary to participate in a study requiring multi-criteria decision-making on a topic requiring no
prior knowledge. Additionally, due to the computer requirements for students in the Lindner College of Business, we could reasonably assume that these subjects would possess the basic computer literacy necessary for navigating our computer-based survey and own the required technology necessary to participate. In compliance with Institutional Review Board (IRB) requirements, students of the PI and the other dissertation committee members were not recruited to avoid any perceived coercion.

As incentive, students were offered a $5 gift card to one of several on-campus restaurants. The amount selected for the incentive was chosen as a value that is high enough to peak interest and be viewed as a “thank you” for their time, without creating any survey response bias. Despite prior research suggesting providing extrinsic motives in direct proportion to how well participants perform increases cognitive effort from study participants (Forsythe et al., 1994; Holt and Laury, 2002), we were not permitted to provide an additional fee based on participant performance. The university IRB stated that incentives must be fairly distributed to all potential participants, and a pay-for-performance incentive could be perceived as coercive to a student who wanted to quit mid-study because he was uncomfortable for some reason. In light of this restriction, we determined *a priori* that we would exclude data for participants whose overall time to complete the experiment is statistically significantly lower than the average time of the participants.

The study was administered in a controlled lab-environment where noise and interruptions were minimized. Subjects arrived to one of the 32, 1-hour time blocks reserved in the lab for this experiment. To ensure consistency, prior to entering the lab, all subjects were provided a simple, yet detailed set of printed instructions (Appendix
A6). For each participant, a link and password for a specific pre-assigned treatment combination (not disclosed to the participants) was provided on their instruction sheet. Since students participated on their own laptops, the passwords were changed immediately after each laboratory session to prevent participants from accessing the study outside of the laboratory. Treatment specifications followed a stratified design to provide coverage along the full-factorial design.

Subjects were seated at one of six long tables, arranged in a U-shape so they had sufficient work space and they were unable to see the computer screens of other test subjects (Figure 5). All subjects wore headphones and were unable to hear the interruption or framing videos presented to other subjects. They were also provided with a pen and plain paper for note taking, if desired.

Figure 5. Arrangement of Subjects in Student Experiment

3.2.3 Crowdsourcing Experiment

The data collection for the crowdsourcing experiment took place over 24-hour period and was conducted via Amazon’s Mechanical Turk. The target subjects (n=108) were MTurkers that met the following qualifications: located in the United States; at least 500 previously approved HITs; and at least a 95% approval rate from prior HITs.
Qualifications are prescreening questions provided by Amazon as a means to assist requestors in filtering out participants. Several filters are provided by and verified by Amazon (such as the three selected for this study), but requestors also have the ability to create custom filters. The United States was chosen as the location in order to increase the likelihood that participants possessed a sufficient grasp of the English language to understand the contents of the study. While participants were also asked if English was their primary language, there is no way to verify their answer to that question and the “location” qualification is verified by Amazon. Requiring study participants to have a minimum number of prior HITs with a pre-specified approval rating are common qualification requirements in studies conducted via Mechanical Turk (Chandler et al., 2014).

While there is much discussion around the best way to incentivize participants on Mechanical Turk, one study from Horton et al. at Harvard found the following: subjects preferred earnings evenly divisible by 5; subjects worked less when pay was lower but they did not work less when the task was more time-consuming; and the median wage was $1.38 per hour (Horton et al., 2010). Participants in our study were offered $0.50 for a task they were informed would take approximately 10-15 minutes, a rate of roughly $2-$3 per hour. While a pay-for-performance incentive was considered, it is known that MTurk participants share information and interact in online communities such as mturkforum.com, Reddit, and Facebook. Thus, it is no surprise that prior studies have demonstrated that participants are more inclined to share information when incentives are offered for correct responses (Edlund et al., 2009).
Because this study could not be conducted in a controlled laboratory experiment, some additional measures were taken to provide us additional information about the participants and their study environment. As suggested by Paolacci et al. (2012), our first attempt to control for quality was to build in measures that ensure independent responses to our four treatments. We created 4 different HITs, one for each treatment, and listed them with the same title but differentiated them by Group A, Group B, Group C, and Group D. (This is a common practice for studies with multiple arms.) Participants were informed up front that HITs would be rejected for anyone who participated more than once or participated in more than one group. We were able to track this information by requiring participants to provide their unique and verifiable Mechanical Turk number. While an individual could feasibly maintain more than one Mechanical Turk account, given Amazon’s registration requirements it is not easily accomplished and not believed to be a prevalent problem.

As suggested by multiple studies, our second attempt to address quality was through the use of an “attention check” (Mason & Suri, 2012; Goodman et al., 2013; Chandler et al., 2014). Participants were informed up front that the HITs would be rejected if participants failed to answer questions that “make sure the participants are paying attention, not providing gibberish for answers, and not speeding through the survey.” On the screen detailing the information about the countries, we included the following question; “If you are paying attention, answer “Strongly Agree” to this question” (Appendix A7). Our third attempt to address quality was to list a set of “requirements” that participants had to “agree” to prior to starting the study: availability of pen and paper; audio/video capabilities; 10-15 minutes of uninterrupted time; and a
request to not discuss the content of our study on public forums. While we acknowledge that there is no way to verify such information, we asked them to agree to this information before participating in the study and concluded with some questions at the end of the study that would not affect their pay for their HIT. We asked them the following questions: to rate on a scale from 0-100 how much effort they put into the survey; rate on a scale from 0-100 the noise level around them while they were taking the study; the type of location where they took the survey; and if they were interrupted while they were taking the survey (Appendix A8). We accepted that we may have to pay participants who were not truthful in their responses to these questions and we paid participants we would have preferred to exclude, we strived to provide incentive for participants to be truthful in their responses by emphasizing that their responses did not affect their pay. Finally, at the end of the survey, we provided the participants an opportunity to share any comments or thoughts about the survey. It was not a question that required a response but allowed us to gather any additional information if they chose to provide it.

As noted earlier, some researchers also suggested employing some form of Frederick’s Cognitive Reflection Test (CRT) to gauge cognitive effort (Goodman et al., 2013). These test are designed to measure a person’s tendency to override their “gut” response and engage in reflecting. For example, you might ask the participant the following: “If a bat and a ball cost $1.10 in total and the bat costs $1.00 more than the ball, how much does the ball cost?” The wrong, but impulsive answer, would be ten cents. The correct answer would in fact be five cents given that $X + (X + 1) = 1.10$, and therefore $X = 0.05$. 
These tests have been found to correlate highly with measures of intelligence (Frederick, 2005). While this may have added value to our study, we had two primary concerns with including a CRT. First, due to information sharing among users, many of participants are familiar with many of the tested and verified CRT questions and answers (Chandler et al., 2014). Thus, the validity of participant responses to these types of tests administered on crowdsourcing sites is questionable. Second, for this particular study, the addition of a cognitively challenging question would only further increase information load and could decrease the amount of focus provided to the details of the actual study. For both of these reasons, we did not include a CRT in our experiment we conducted on Mechanical Turk.

3.3 Results and Conclusions

3.3.1 Student Experiment

In total, n=28 students participated in the pilot experiment. Data were excluded for n=3 participants due to technical issues while participating in the study: one subject’s computer battery died; one subject’s computer froze; and a technical glitch in the Mechanical Turk software led to the responses of one participant to stop recording mid-experiment\(^3\). Additionally, data were excluded for n=6 participants for lack of sincere effort (integrity). Comparing the overall time to complete the experiment for the n=6 participants without sincere effort vs. the remaining participants was as follows: Minimum time to complete the study was 136 sec vs. 465 sec; median time to

\(^3\) On the day that the glitch occurred, the PI observed this participant complete the experiment. As this subject was the only participant on that day, we were confident that the experiment was completed but unsure as to why Mechanical Turk only recorded a fraction of the data and marked the participant’s responses as incomplete.
complete the study was 242.5 sec vs 755 sec; max time to complete the study was 391 sec vs. 1405 sec (Figure 6).

**Figure 6: Time to complete study**

![Time to Complete Survey](image)

After excluding the n=9 data points as explained above, there were n=19 participants in the study. There were n=5 participants in treatments A (no interruption & no emphasis frame), B (interruption & no emphasis frame), and D (interruption & emphasis frame) and n=4 participants in treatment C (no interruption & emphasis frame). Overall, there were more women than men (Table 2).

### Table 2. Study subjects

<table>
<thead>
<tr>
<th>Group</th>
<th>Technical Issues</th>
<th>Integrity</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (INT, EF)</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>B (INT, EF)</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>C (INT, EF)</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>D (INT, EF)</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Grand Total</td>
<td>3</td>
<td>6</td>
<td>13</td>
<td>6</td>
<td>28</td>
</tr>
</tbody>
</table>

In order to evaluate the timeliness of decision-making, we measured the overall time to complete the study. We hypothesized that interruptions increase decision-making time (H1) and an emphasis frame reduces decision-making time (H3). Table 3
and Figure 7 illustrate the minimum, mean, median, maximum, and standard deviation of the time to complete the study by treatment group.

In order to evaluate the quality of decision-making, we calculated their total score based on the previously described scoring rubric. We hypothesized that interruptions decreased decision-making quality (H2) and an emphasis frame improved decision-making quality (H4). Table 4 and Figure 8 illustrates the minimum, mean, median, maximum, and standard deviation of the score by treatment group. Due to the resulting small n, we were unable to attain statistically significant results.

Table 3: Time to complete student experiment (time in sec)

<table>
<thead>
<tr>
<th>Group</th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (INT, EF)</td>
<td>488.9</td>
<td>774.3</td>
<td>521.2</td>
<td>1,359.8</td>
<td>389.0</td>
</tr>
<tr>
<td>B (INT, EF)</td>
<td>668.4</td>
<td>866.6</td>
<td>892.5</td>
<td>1,084.9</td>
<td>187.9</td>
</tr>
<tr>
<td>C (INT, EF)</td>
<td>487.5</td>
<td>674.0</td>
<td>679.6</td>
<td>849.2</td>
<td>175.9</td>
</tr>
<tr>
<td>D (INT, EF)</td>
<td>380.3</td>
<td>566.8</td>
<td>603.9</td>
<td>637.9</td>
<td>105.9</td>
</tr>
</tbody>
</table>

Figure 7: Time to complete student experiment
### 3.3.2 Crowdsourcing Experiment

In total, n=147 MTurkers participated in the pilot experiment, but only n=108 completed the study. The completion and dropout rates were distributed as follows by treatment group: A n=27 (n=13 incomplete); B n=28 (n=13 incomplete); C n=26 (n=7 incomplete); D n=27 (n=6 incomplete) (Table 5). Of the n=108 participants who completed the study, there were n=61 female participants (56.48%) and n=47 male participants (43.52%) (Table 6). They ranged in age from 22 years old up to 65 years old with the 25-34 age range being the most common (Table 7). Participants (n=45)
reported that their highest level of education was a bachelor’s degree more often than any other response, but there was a wide variety of responses reported (Table 8). All n=108 participants reported that English was their primary language and n=2 participants reported being interrupted while participating in the study and both were in treatment group B.

Table 5: Completion and Dropout Rates from Mechanical Turk Experiment

<table>
<thead>
<tr>
<th>Group</th>
<th>Complete</th>
<th>Dropout</th>
<th>Total</th>
<th>% Dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (INT, EF)</td>
<td>27</td>
<td>13</td>
<td>40</td>
<td>32.5%</td>
</tr>
<tr>
<td>B (INT, EF)</td>
<td>28</td>
<td>13</td>
<td>41</td>
<td>31.7%</td>
</tr>
<tr>
<td>C (INT, EF)</td>
<td>26</td>
<td>7</td>
<td>33</td>
<td>21.2%</td>
</tr>
<tr>
<td>D (INT, EF)</td>
<td>27</td>
<td>6</td>
<td>33</td>
<td>18.2%</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>39</td>
<td>147</td>
<td>26.5%</td>
</tr>
</tbody>
</table>

Table 6: Reported Gender from Mechanical Turk Experiment

<table>
<thead>
<tr>
<th>Group</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>A (INT, EF)</td>
<td>16</td>
<td>11</td>
<td>26.2%</td>
</tr>
<tr>
<td>B (INT, EF)</td>
<td>16</td>
<td>12</td>
<td>26.2%</td>
</tr>
<tr>
<td>C (INT, EF)</td>
<td>16</td>
<td>10</td>
<td>26.2%</td>
</tr>
<tr>
<td>D (INT, EF)</td>
<td>13</td>
<td>14</td>
<td>21.3%</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>47</td>
<td>56.5%</td>
</tr>
</tbody>
</table>

Table 7: Reported Ages from Mechanical Turk Experiment

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Treatment Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>&lt;25</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>25-34</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>35-44</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>45-54</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>55+</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>
Table 8: Reported Education from Mechanical Turk Experiment

<table>
<thead>
<tr>
<th>Reported Highest Level of Education</th>
<th>Treatment Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some high school</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>High school graduate</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Some college credit, but no degree</td>
<td></td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>12</td>
<td>31</td>
</tr>
<tr>
<td>Associate's degree</td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Bachelor's degree</td>
<td></td>
<td>11</td>
<td>11</td>
<td>15</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>Master's degree</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Professional degree (e.g. - MD, DDS, DVM, JD)</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doctorate degree (e.g. - PhD, DBA, EdD)</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>27</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>108</td>
</tr>
</tbody>
</table>

As noted earlier, the study design included some questions to help us gauge the quality of the responses provided. First, in response to our “attention check” built into the middle of the study (see Appendix A7), we found that 72.2% of the participants (n=78) provided the correct response to the question, 2.8% of the participants (n=3) participants answered it incorrectly, and 25% (n=27) of the subjects completely missed answering the question at all (Table 9).

Table 9: Accuracy of “Attention Check” from Mechanical Turk Experiment

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Attention Check</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correctly Answered</td>
<td>%</td>
<td>Incorrectly Answered</td>
<td>%</td>
<td>Missed</td>
<td>%</td>
</tr>
<tr>
<td>A (INT, EF)</td>
<td>18</td>
<td>66.7%</td>
<td>2</td>
<td>7.4%</td>
<td>7</td>
<td>25.9%</td>
</tr>
<tr>
<td>B (INT, EF)</td>
<td>20</td>
<td>71.4%</td>
<td>0</td>
<td>0.0%</td>
<td>8</td>
<td>28.6%</td>
</tr>
<tr>
<td>C (INT, EF)</td>
<td>19</td>
<td>73.1%</td>
<td>1</td>
<td>3.8%</td>
<td>6</td>
<td>23.1%</td>
</tr>
<tr>
<td>D (INT, EF)</td>
<td>21</td>
<td>77.8%</td>
<td>0</td>
<td>0.0%</td>
<td>6</td>
<td>22.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>78</td>
<td>72.2%</td>
<td>3</td>
<td>2.8%</td>
<td>27</td>
<td>25.0%</td>
</tr>
</tbody>
</table>

Second, we requested that the participants self-report on a scale from 0-100, the level of effort they put into the study. Their responses ranged from 20% effort on the low end up to 100% on the high end with exactly half of the participants reported they gave an effort level of 95% or greater (Table 10). When looking at the values from these two
questions together, we found that 24.1% (n=13) participants who reported an effort of 95% or greater missed the “attention check” question (Table 11, Figure 9).

**Table 10: Self-Reported Study Effort from Mechanical Turk Experiment**

<table>
<thead>
<tr>
<th>Percent Effort</th>
<th>Treatment Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-50</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>51-75</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>76-85</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>86-94</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>95+</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

**Table 11: Self-Reported Effort vs. Attention Check from Mechanical Turk Experiment**

<table>
<thead>
<tr>
<th>% Effort</th>
<th>Incorrect/Missed</th>
<th>Correct</th>
<th>Total</th>
<th>% Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-75</td>
<td>6</td>
<td>14</td>
<td>20</td>
<td>30.0%</td>
</tr>
<tr>
<td>76-85</td>
<td>6</td>
<td>11</td>
<td>17</td>
<td>35.3%</td>
</tr>
<tr>
<td>86-94</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>29.4%</td>
</tr>
<tr>
<td>95+</td>
<td>13</td>
<td>41</td>
<td>54</td>
<td>24.1%</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>78</td>
<td>108</td>
<td>27.8%</td>
</tr>
</tbody>
</table>

**Figure 9: Self-Reported Effort vs. Attention Check from Mechanical Turk Experiment**
Third, we compared the total time to complete the experiment to their self-reported effort (Table 12, Figure 10). While the average time to complete the study did go up as the percentage of self-reported effort went up, the two variables were weakly correlated, $r(108) = 0.173$, $p<0.037$.

**Table 12: Time to Complete Experiment by Self-Reported Effort from Mechanical Turk Experiment**

<table>
<thead>
<tr>
<th>% Effort</th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-50</td>
<td>421</td>
<td>509</td>
<td>509</td>
<td>597</td>
<td>124</td>
</tr>
<tr>
<td>51-75</td>
<td>295</td>
<td>551</td>
<td>489</td>
<td>1,348</td>
<td>248</td>
</tr>
<tr>
<td>76-85</td>
<td>297</td>
<td>576</td>
<td>537</td>
<td>1,453</td>
<td>268</td>
</tr>
<tr>
<td>86-94</td>
<td>132</td>
<td>629</td>
<td>565</td>
<td>1,200</td>
<td>336</td>
</tr>
<tr>
<td>95+</td>
<td>220</td>
<td>690</td>
<td>617</td>
<td>1,997</td>
<td>367</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>132</strong></td>
<td><strong>636</strong></td>
<td><strong>597</strong></td>
<td><strong>1,997</strong></td>
<td><strong>328</strong></td>
</tr>
</tbody>
</table>

**Figure 10: Time to Complete Experiment by Self-Reported Effort from Mechanical Turk Experiment**

% Effort vs. Total Time
Finally, we were unable to conduct an analysis of variance (ANOVA) with fixed factors for interruptions or the attention check due to inequality of variance. Given the unequal sample sizes, the results of Levene’s test for homogeneity of variance (Table 13) suggests that any ANOVA inferences could be affected and a post-hoc test for comparison between means was not conducted. The assumptions for homogeneity of variance was upheld for our emphasis framing results but the results were not statistically significant for either of our main effects; Decision-Making Time $F(1,106) = 0.191, p = 0.663$; Quality Score $F(1,106) = 3.414, p= 0.067$.

**Table 13: Levene’s Test of Equality of Error Variances**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision-Making Time</td>
<td>Attention Check</td>
<td>4.283</td>
<td>1</td>
<td>106</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>Interruption</td>
<td>6.081</td>
<td>1</td>
<td>106</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Emphasis Frame</td>
<td>1.358</td>
<td>1</td>
<td>106</td>
<td>0.247</td>
</tr>
<tr>
<td>Quality Score</td>
<td>Attention Check</td>
<td>4.822</td>
<td>1</td>
<td>106</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>Interruption</td>
<td>0.855</td>
<td>1</td>
<td>106</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>Emphasis Frame</td>
<td>0.695</td>
<td>1</td>
<td>106</td>
<td>0.406</td>
</tr>
</tbody>
</table>

Based on all of these results combined, there was significant concern with quality of data. In an effort to ensure the integrity of our study, it was determined that the data from Mechanical Turk should be discarded.

In light of our decision about the Mechanical Turk data, we sought to evaluate the responses to the voluntary question at the end of our experiment, which provided participants the opportunity to share any comments or thoughts on our study. The vast majority of participants ($n=82$) did not have any comments. But, there were $n=6$ participants who indicated they felt that the study induced information overload. Given the objective of our study, while some comments expressed frustration, it simply provided further evidence for the appropriateness of our design. Some of the comments
were as follows: “too much information”; “Data overload”; “That felt like an overwhelming amount of information to assess and process.” In terms of negative feedback not related to the concern with information overload, n=2 participants expressed disappointment with their financial incentive: “fairly low pay considering the effort required”; “too much information to compile and absorb for the pay”.

Finally, just under 10% of the participants shared comments that illustrated they found our study interesting or expressed enjoyment in their participation. Some examples of these comments are as follows: “Interesting and thought provoking”; “great survey! very creative and engaging.”; “It was a good engaging study enjoyed it much and thanks.”; “was fun thank you!”. These results seemed to support the notion suggested in a study conducted by Kaufmann et al. (2011) that suggested Mechanical Turk workers were not only driven by extrinsic financial incentives but at least partially intrinsically motivated by the fact that they find participating in HITs to be fun or enjoyable.

3.4 Discussion

The results of these studies highlighted some of the challenges facing researchers interested in developing the science behind knowledge-work and removing the barriers to knowledge-worker productivity. Studying the complex cognitive tasks synonymous with knowledge work is difficult. Despite our challenges, we still feel empirical methods are a natural fit for the research, specifically, experiments that are particularly well suited to the task given their strength for precision and control. Though they offer less external validity than other empirical methods, a well-designed laboratory experiment
can help parse out the complex tasks associated with knowledge work while abstracting away unnecessary details without a loss of generalization. But, we feel that our results highlight that while subject recruitment is a nontrivial decision in any experiment, it is considerably more important when conducting experiments to develop the science behind knowledge-worker productivity. Sincere cognitive effort on behalf of the participants is critical to the integrity of the results.

Ensuring sincere cognitive engagement of participants is a challenging but important topic when conducting these studies. Prior research in operations, psychology, behavioral economics, and neuroscience suggests that motivation plays an important role in cognitive control because individuals must be sufficiently motivated to make the conscious choice to engage in a cognitive task. Current literature from a variety of fields suggest that given properly aligned incentives (intrinsic and/or extrinsic), individuals can be motivated to provide a sufficient level of cognitive effort. The challenge in conducting experiments is determining which incentive works best for knowledge-intensive workers and which incentives work best in the environment being studied.

The operations literature has published evidence of intrinsically and extrinsically motivating students in laboratory experiments. The general consensus among research on extrinsic motivation (i.e. – cash, course credit) illustrated that cognitive effort of study participants increased when given in direct proportion to how well participants perform (Schweitzer & Cachon 2000; Bolton & Katok, 2008; Bostian et al., 2008). Alternatively, other researchers found success with intrinsically motivating cognitive effort (Lepper et al. 1973; Read, 2005). Despite the fact that students regularly participate in knowledge-intensive tasks as they complete their coursework, we struggled with gaining sincere
cognitive engagement. In our study, we relied on a nominal extrinsic reward to encourage students to take part in our study and relied on intrinsic motivation to perform the knowledge intensive task. We did not find this successful and suggest working closely with IRB to establish a means for offering a pay-for-performance incentive in light of the literature that suggests the benefits of this reward structure.

In terms of our results from the student experiment, while our results lacked statistical significance, they do suggest potential given a larger sample size. While the authors are not aware of any prior studies investigating emphasis framing as an operational tactic to help mitigate the effects of information overload and interruptions on the quality and efficiency of decision-making, our study suggests it may be worth further investigating a relationship between emphasis framing and the quality and efficiency of decision-making.

In our other experiment, conducted with Amazon’s Mechanical Turk, we were able to execute our study significantly faster and more cost effectively than with our student experiment. In general, our results raised significant concerns about the quality of data we collected. There appeared to be a lack of sincere cognitive effort across the board. Aside from general concerns a researcher may have about lack of control for the environment (noise, interruptions, etc…), the research on motivation suggests we further examine whether the intrinsic or extrinsic motivation was sufficient for the knowledge-intensive task presented to the participants.

In terms of the research on extrinsic motivation, while several studies found evidence illustrating that the cognitive effort of study participants increased when given in direct proportion to how well participants performed, their research did not take into
account some of the unique challenges introduced when participants on crowdsourcing sites such as Amazon’s Mechanical Turk share information and interact in online communities such as mturkforum.com, Reddit, and Facebook. While results from an experiment conducted on a crowdsourcing site might demonstrate improved performance with a pay-for-performance incentive, the results might include a significant bias.

Prior studies have demonstrated that participants are more inclined to share information when incentives are offered for correct responses (Edlund et al., 2009). In an attempt to avoid this bias, we offered a flat monetary incentive of roughly $2-$3 per hour that was above the $1.38 median hourly reservation wage reported as the standard in studies conducted on Mechanical Turk (Horton et al., 2010), and we still encountered significant data quality issues. While a question may arise as to whether the pay was sufficient enough to motivate high quality work, at least two different studies demonstrated no effect on the quality of work as the wage earned on Mechanical Turk was manipulated (Mason & Watts, 2009; Marge et al., 2010). Mason and Suri (2011) advise paying less than the expected reservation wage and then increasing the wage if the completion rate is too late.

We are comfortable with the payment aligned with our experiment for three reasons. First, n=108 participants completed our experiment in less than 24 hours. Second, there was only one comment out of all n=108 participants that openly expressed disappointment with the pay. While we recognize that not all workers will inform a requestor when they feel the pay is too low, it has been noted by other researchers that the rich online community that has developed around MTurk is very active and workers
often share bad experiences (e.g., when relative pay is egregious) in online forums (Paolacci et al., 2010; Mason & Suri, 2012). To further research the “fairness of pay”, the PI did locate \( n=10 \) negative comments posted in the online “turker” communities about this study. Of the ten comments, nine of them highlighted their frustration with being informed that their HIT was rejected for failing to correctly respond to the attention check. For example, “I’ve done 700+ mTurk surveys and never missed an attention check before.” Thus, given the overall concern with their “rejected HIT”, these comments provided us less of an insight into the fairness of our incentive and more insight into the complex dynamics between workers and requestors, especially when considering rejected HITs. Given the challenges posed by pay-for-performance and the less than impressive results with flat fees as an extrinsic reward we encountered in our study, we suggest that extrinsically motivating crowdsourcing participants to engage in cognitively challenging tasks may be extremely difficult to achieve at this time.

Given the challenges with extrinsically motivating participants, the option for trying to intrinsically motivate participants is another alternative. While gauging the intrinsic motivation of participants in any study is difficult, our voluntary question at the end of our experiment, which provided participants the opportunity to share any comments or thoughts on our study, at least provided us with some potentially interesting insight. Nearly 10% of our participants voluntarily provided positive feedback suggesting some level of enjoyment from participating in our study. Thus, these results suggest that even intrinsic motivation may be insufficient to motivate significant cognitive effort from crowdsourcing participants.
Finally, if it is challenging to effectively motivate crowdsourcing workers intrinsically or extrinsically, Mechanical Turk has a built-in reward vs. punishment incentive system. Requestors have the ability to “reject” a HIT. As noted, in our study, we informed participants up front that a participant check was built into the study and, in the event that they failed to correctly answer the question, their HIT would be rejected (i.e. – they not only would not get paid but it would adversely influence the workers “ratings”). In theory, the fear of a “punishment” from a rejected HIT should increase the quality of work. While we do not intend to have a full discussion of the implications of this built-in feature, in light of the fact that a worker’s “reputation” and pay is directly influenced by a requestor that rejects a HIT, we suggest that this is not a simple “reward vs. punishment” feature. Due to the risks to the worker, very active Mechanical Turk workers regularly communicate with each other in a variety of online communities such as Facebook, Reddit, and Turker Nation to provide positive and negative feedback around specific HITs or requestors. While it’s infeasible (and likely impossible) for researchers to locate all comments posted about their study, some people have suggested that the more popular sites can have a strong effect on the acceptance rate of HITs (Mason & Suri, 2012). Given the difficulty of cognitively challenging tasks synonymous with knowledge work, requestors need to be aware that rejecting HITs for poor quality, as occurred in our study, could lead workers to create a stream of negative feedback and taint the sample.

Overall, given the results of our experiment and the challenges with establishing appropriate motivation for participants to engage sufficient cognitive effort for knowledge intensive tasks, we suggest that Mechanical Turk may not be suitable for research.
designed to develop a greater understanding of the science behind knowledge work at this time. Further research is needed to develop a better understanding of the unique challenges introduced by this new platform and how these factors interact with each other.
The challenges of studying the complex and cognitively challenging tasks associated with knowledge work

ABSTRACT

Technological advances, starting with the industrial revolution, have led to the disappearance of blue collar jobs and a manufacturing-based economy in favor of a knowledge-based economy where knowledge workers drive productivity and economic growth. In this new environment, investigating operational problems through a behavioral lens is becoming increasingly more essential and the field of operations has seen the emergence and growth of Behavioral Operations Management (BOM). Humans are central to many aspects of operations management (OM) and successful implementation of OM tools and techniques in knowledge-work settings relies on

4 This chapter is written as a stand-alone paper that will be submitted to the Journal of Operations Management.
understanding the interaction between human behavior and the operational systems we strive to improve. Developing the science behind knowledge-work, removing the barriers to knowledge-worker productivity, and investigating cognitively challenging tasks is difficult. Given the advantage of isolating specific effects, greater use of laboratory experiments are needed to study these complex processes. And while this is a noteworthy methodology used for decades in the psychology, economics, and behavioral decision research literature, it is among one of the least used in the operations literature. This paper examines the emergence of research dedicated to knowledge-work, discusses some of the challenges of studying the complex cognitive tasks synonymous with knowledge work, and provides some general recommendations for those considering laboratory experiments involving knowledge workers.

4.1 Introduction

According to Peter Drucker (1999), “the most valuable asset of a 21st-century institution, whether business or non-business, will be its knowledge workers and their productivity.” Knowledge workers currently outnumber manual workers 2-to-1 in the United States (Bureau of Labor Statistics) and growth is expected to continue in these occupations for the foreseeable future. While definitions vary slightly from source to source, knowledge workers have been described as employees with a formal education or high degree of expertise in a particular area, who leverage their knowledge and intellectual capacity to transform information into some form of “product” (Drucker, 1999; Davis, 1999; Davenport et al., 2002; Ramirez & Nembhard, 2004; Hopp et al., 2009). Knowledge work is inherently more cognitive than physical in nature, and according to
Hopp et al. (2009), “knowledge work is considered a subset of white-collar work, because highly knowledge-intensive tasks are classified as white collar.” Examples include business and financial operations occupations such as analysts or accountants, general management occupations, computer and mathematical occupations such as programmers and actuaries, legislators, medical doctors, lawyers, and scientists. There is significant need to shift from the standard throughput focus that was necessary in the manufacturing-based economy of the 20th-century and focus on developing the science behind knowledge-work. It is essential that we discover ways to remove the barriers to efficient and high-quality knowledge-work (Drucker, 1999; Davenport et al., 2002; Ramirez & Nembhard, 2004; Hopp et al., 2009; Froehle & White, 2014).

While a relatively limited amount of research on the productivity of knowledge work has been conducted to date, the growing field of behavioral operations management (BOM) has established a strong foundation for conducting research on this important topic. According to Gino and Pisano (2008), “behavioral operations explores the intersection between behavioral decision research, which is focused on human behavior, and operations management which is focused on system behavior.” As behavioral operations has evolved, Croson et al. (2013) further clarified this definition by highlighting the micro-level focus of the field stating “Research in behavioral operations analyzes decisions, the behavior of individuals, or small groups of individuals.” While macro-level issues, such as organizational learning are important topics of research, they emphasized the importance of limiting the scope of BOM research to understanding the role of the individuals and small groups and how they influence
operations systems and processes. Research focusing on the productivity of knowledge-workers fits well within the scope of behavioral operations management.

The emergence and growth of the field of behavioral operations management began in the late 1980’s and early 1990’s when several journal articles highlighted the importance of and need for more empirical work in OM (Adam & Swaimidass, 1989; Meredith et al., 1989; Flynn et al., 1990). Empirical research is the systematic process of deriving and analyzing data from direct or indirect observations in order to develop and test theories about the operating processes and systems the field strives to improve (Roth, 2007). While empirical methods have been widely applied across many areas within operations management, empirical methods are a natural fit for behavioral operations management research. They bridge the gap between analytical models and real world business problems by helping to develop an understanding of the influence of human behavior on how operating systems work, perform, and respond to management interventions (Gino & Pisano, 2008).

This early focus on empirical research laid the foundation for the significant growth in BOM research that followed the call for incorporating behavioral factors in OM by Boudreau et al. (2003). Behavioral operations management is now an accepted sub-field in the discipline of OM as highlighted by a variety of significant developments in the last decade: the commencement of an annual BOM conference starting in 2006; an INFORMS Section of Behavioral Operations Management; a POMS College of Behavior in Operations Management; and three special issues in the last decade in two top OM journals. In 2006, the Journal of Operations Management (JOM) published the first special issue dedicated to the topic (Boyer & Swink, 2006). The sixteen articles selected
for that pioneering issue were important for providing a framework to identify behavioral assumptions commonly used in OM analytical models and identifying OM problems that could be better explained or investigated through a behavioral lens. In 2008, the *Manufacturing and Service Operations Management Journal (MSOM)* published a behavioral operations management special issue with seven publications that highlighted some of the early findings from research in this emerging field (Gans & Croson, 2008). Most recently in 2013, *JOM* again dedicated a special issue (Croson et al.) that was instrumental in more concretely defining the scope of behavioral operations management and highlighting the expanding contextual and methodological diversity of BOM research. All of this early research has been important for helping the field identify key OM problems that could benefit from investigation through a behavioral lens, defining and outlining a potential research agenda, and illustrating some early examples of rigorous behavioral operations research. Despite this recent growth in BOM contributions, research focused on the productivity of knowledge workers has been sparse.

4.2 Background

4.2.1 Knowledge-Work

The fields of operations research and management science contributed to a greater than 50-fold increase in the productivity of manual workers over the last century (Drucker, 1999). But, there has been a significant decline in manual workforce and as the distribution of the labor force has shifted, there is significant need to further develop the science behind knowledge-work and remove the barriers to efficient and high-quality
knowledge-work (Drucker, 1999; Davenport et al., 2002; Ramirez & Nembhard, 2004; Hopp et al., 2009; Froehle & White, 2014). But, it is challenging to measure the productivity of the complex cognitive tasks synonymous with knowledge work.

In order to begin developing this field, Ray and Sahu (1989) helped by defining the types of cognitive tasks often associated with knowledge work. They classify knowledge-intensive tasks into two categories based on their overall complexity and required mental effort. In the first category are routine or repetitive tasks requiring minimal mental effort, such as clinical procedural tasks. The inherent nature of these tasks places limited cognitive demand on the worker, but they are still classified as knowledge-intensive due to a specific level of expertise or education that are required for their completion. In the second category are non-repetitive or non-routine tasks that place greater cognitive demand on the worker due to the fact that the tasks generally involve the processing and synthesis of complex, interrelated information. A defining attribute of knowledge workers is their aptitude for the non-repetitive, non-routine cognitive tasks and their ability to generate knowledge and make decisions.

Daniel Kahneman’s (2011) work on the cognitive process and decision-making has a different theoretical basis than the research in the operations literature, but similar in that he postulates that the brain uses two fundamentally different systems to process information and make decisions. He posits that activation of System 1 occurs when workers are presented with routine and repetitive tasks, such as clinical procedural tasks mentioned above, and decisions are made quickly and intuitively. When an individual is presented with non-routine or non-repetitive tasks, information-intensive tasks, System 2 is activated and decisions are slower, more deliberate, and often
require intense focusing. Additionally, the tasks that require the activation of System 2 require significant cognitive exertion and the decision to activate System 2 must be a deliberate action taken up by the individual. The focus of this discussion is focused on System 2 decision-making that is a necessary for many knowledge-intensive tasks.

Given the types of tasks associated with knowledge work, Ramirez and Nembhard (2004) identified ways of measuring knowledge-worker productivity. They emphasize the importance of quantifiable productivity measures that differ from the commonly used throughput measures used in manufacturing that acknowledge the uniqueness of knowledge workers, and that account for the different dimensions of performance. Additionally, instead of a single measure of knowledge-worker productivity they advocate for some combination of the following: quality, cost and/or profitability, timeliness, autonomy, efficiency, quality, effectiveness, customer satisfaction, innovation/creativity, project success, responsibility/importance of work, knowledge worker’s perception of productivity, and absenteeism. We suggest that for now, two of the most important measures of productivity are the quality and timeliness of task completion. As quality is central to most aspects of OM, it is also one of the most important measures of productivity among knowledge workers. Additionally, in an era when Americans take in five times more information every day than they did 30 years ago (Levitin, 2014), the ability to process the deluge of information in a timely manner is becoming an increasingly important measure of productivity among knowledge workers.

Knowledge workers leverage their knowledge and intellectual capacity to transform information into some form of “product.” The generation of this “product” relies processing of information and ultimately the decision-making of the knowledge workers.
Prior research in psychology and economics have shown that humans are inherently irrational and human behavior and cognition affects decision-making (Tversky & Kahneman, 1974; Simon, 1979). In other words, humans’ actions are influenced by their beliefs, biases, motivations, and cognitive limits and often deviate from logically optimal solutions even when presented with perfect (full and accurate) information. These findings are in distinct contrast to traditional OM models, which often assume rational behavior of decision makers, such that all decisions will serve to maximize some sort of utility function. This contradiction highlights the importance and need for incorporating behavioral factors into OM and use of empirical methods to help develop the science behind knowledge-worker productivity.

In the last couple of decades, not only has the number of empirical BOM papers continued to rise, but so has the contextual and methodological diversity. From a contextual perspective, published empirical work in BOM has most commonly addressed inventory, production, and supply chain issues (Bendoly et al., 2006; Croson et al., 2013). Despite this recent growth in the empirical behavioral operations literature, from a contextual perspective, research focused on productivity of knowledge work has largely been underrepresented. The literature that has been published can be categorized into two groups: 1) knowledge tasks that rely on information being processed and synthesized by an individual knowledge worker and, 2) knowledge tasks that require the collective collaboration of a group of knowledge workers.

At the most basic level, knowledge work is conducted on the individual level. To date, research concerned with the productivity of individual knowledge tasks has primarily focused on how systems and processes are influenced by the individual
biases, motivations, and cognitive limits of the workers themselves. One of the earliest
test
examples was Schweitzer and Cachon (2000), which empirically examined individual
decision bias in inventory management. Different variations of the newsvendor problem
have since been explored in several papers, continuing the focus on different aspects of
individual decision-making bias (Bolton & Katok, 2008; Bostian et al., 2008). In research
not related to the newsvendor problem, Bendoly (2011) explored links between
individual motivation and optimal decision-making in revenue management. And finally,
research from Zellmer-Bruhn (2003) and Froehle and White (2014), showed that the
limited cognitive capacity of knowledge workers can influence the productivity of
knowledge work by illustrating how interruptions and forgetting can influence task
performance.

Many organizations form teams of employees to complete knowledge tasks that are
highly collaborative in nature, and their productivity is highly dependent on the
interaction among a group of knowledge workers (e.g., brainstorming solutions to a
problem). Research in this area has focused on developing an understanding of the
influence of incentives (Oliva & Watson, 2009; Katok & Siemsen, 2011), trust (Amaral &
Tsay, 2009; Park & Keil, 2009), knowledge acquisition (Cummings, 2004), and
collaboration (Bendoly et al., 2008; Girotra et al., 2010) on the productivity of knowledge
work.

From a methodological perspective, published empirical work in BOM has been
conducted using a wide variety of qualitative and quantitative approaches. The most
common methods used have been surveys, case studies, and experiments (Roth, 2007;
Boyer & Swink, 2008; Croson et al., 2013). While many different empirical methods may
be appropriate for investigating knowledge-worker productivity, laboratory experiments are particularly well suited to the task given their strength for control. Though they offer less external validity than other empirical methods, a well-designed laboratory experiment can help parse out the complex tasks associated with knowledge work while abstracting away unnecessary details without a loss of generalization. Despite its apparent advantages for this type of research, use of laboratory experiments to develop the science behind knowledge work has been minimal. We theorize that this apparent gap in the literature is due to some of the challenges associated with conducting this type of research to study the complex cognitive tasks synonymous with knowledge work. But, well-designed experiments that are thoroughly grounded in the literature and have a rigorous approach to the data analysis are much needed and have the potential to significantly influence the goal of developing the science behind knowledge worker productivity.

4.2.2 Experiments with Knowledge Workers

Laboratory experiments are a noteworthy methodology that has been used for years in psychology, economics, and behavioral decision research, but it is among one of the least used among the large body of operations research, which relies heavily on analytical models (Bendoly et al., 2006; Fisher, 2007; Boyer & Swink, 2008). While BOM is still a relatively niche area in operations management research as a whole, the field is rapidly growing and maturing as an accepted sub-field in OM. In a review of behavioral research using controlled experiments in the 20 years from 1985 to 2005, Bendoly et al. (2006) found some interesting publishing trends in the top five journals in
our field [i.e. Production and Operations Management (POM), Journal of Operations Management (JOM), Manufacturing and Service Operations Management (MSOM), Management Science (MS) Decisions Sciences Journal (DSJ)]. First, the number of BOM publications was significantly less in the 1985-1989 time frame than the remainder of that time period. Second, only 19 BOM papers appeared in those top five journals over that entire 20-year time period. Last, not a single article of this nature appeared in MSOM.

With the singular focus on laboratory experiments in the 2006 literature review provided by Bendoly et al., the field of behavioral operations management began to be equated with only experimental research (Croson et al., 2013). In the 2013 survey of the literature, Croson et al. provided evidence that refuted that perception. Of the 100 behavioral operations management articles published from 2006 to 2011 in the same five journals mentioned above, less than half were laboratory experiments and only a fraction of those were related to knowledge-worker productivity. Given that those 100 publications make up less than 5% of the total publications in those five journals over that 6-year time period, and laboratory experiments focused on knowledge-worker productivity make up less than half of those 100 publications, this highlights the simple fact that research in this area is still highly underrepresented in the operations literature.

As the interest in behavioral operations research has expanded, several papers have emerged to provide potential researchers a guide to conducting laboratory experiments (Katok, 2011; Knemeyer & Naylor, 2011; Bendoly & Eckerd, 2013). While we will not provide a full summary of their work, this paper extends their work by providing a brief overview and discussing some unique challenges introduced when
studying the complex cognitive tasks synonymous with knowledge work. In laboratory experiments, a researcher systematically manipulates independent variables and evaluates their effects on specified dependent variables. For this reason, they have broad appeal for their internal validity and are useful in building and verifying theory. High-quality laboratory experiments generally follow the scientific method (Meredith, 1998), be thoroughly grounded in the literature, be well-designed, and have a rigorous approach to the data analysis. In the tradition of an inductive approach to theory-building, the first step to high-quality empirical research is observing and describing a phenomenon of interest (Churchill, 1979; Roth 2007). Next, explanations of relationships among constructs are described and theoretically supported hypotheses are formed (Edwards & Bagozzi, 2000). Finally, the hypotheses are methodologically and rigorously tested and refinements and modifications to theory are made where appropriate (Roth, 2007). While the links between the theory development in the first two steps and the experimental design are inextricable (Kagel & Roth, 1995), we would like to focus our discussion around the unique challenges introduced by the knowledge workers themselves. While selection of subjects to recruit is a nontrivial decision in any experiment, we suggest it is considerably more important when conducting laboratory experiments to develop the science behind knowledge-worker productivity.

As we have already noted, knowledge workers use their education and intellectual capacity to transform information into some form of “product” and are valued for their individual decision-making capabilities. So who are the best subjects for these types of experiments? The three types of participants that are the most common for laboratory experiments are students, participants from the field of study (practitioners), or
participants from the newly available crowdsourcing sites such as Amazon’s Mechanical Turk (MTurk).

Ensuring sincere cognitive engagement of participants is a challenging but important topic when conducting these studies. Motivation plays an important role in cognitive control (Botvinick & Braver, 2015) because individuals must consciously make a choice to exert a sufficient cognitive effort. A surge of new research in psychology, behavioral economics, and, especially, neuroscience has emerged with a focus on understanding the mechanisms by which motivation and cognitive control interact. One group of researchers has suggested that choice is often motivated by a classic reward vs. punishment perspective – a desired outcome that can be brought about by focusing on the cognitively challenging task at hand or with the desire to avoid the outcome that follows from failure to provide sufficient effort (Locke & Braver 2008; Engelmann et al. 2009; Jimura et al. 2010; Savine et al. 2010; Braem et al. 2012). A slightly different perspective, presented in recent research from Shenhav et al. (2013), suggests a cost-benefit model might better explain how motivation and cognitive control interact. Their theory of Expected Value of Control (EVC) proposes that decisions about the level of cognitive effort an individual decides to exert is selected based on the expected cumulative reward from their cognitive exertion less the expected costs of the exertion. Whether the reward or benefit is intrinsic or extrinsic, the level of cognitive control exerted by an individual has been linked to motivation to attain a properly aligned incentive.

While neither as extensive nor as thorough as what has been published in the psychology and neuroscience literature, research from the fields of behavioral
economics and operations research has also touched on the subject of motivation, intrinsic and extrinsic, and incentives in experiments. There have been several studies that have shown how extrinsic motivation (e.g., money, course credit), when given in direct proportion to how well participants perform, increases cognitive effort from study participants (Forsythe et al., 1994; Holt & Laury, 2002). An example would be the pay-for-performance structure used with laboratory experiments exploring the newsvendor problem (Schweitzer & Cachon 2000; Bolton & Katok, 2008; Bostian et al., 2008). Other research has also touched on the notion that intrinsic motivation (e.g., an internal desire to do well), can also be a powerful motive (Lepper et al. 1973; Read, 2005). For example, Samuelson and Bazerman (1985) found that in their challenging “acquire-a-company” problem, they found that financial incentives had little effect (i.e., did not improve performance) because the problem was extremely difficult and the subjects needed to be intrinsically motivated to perform well and financial incentives added little extra motivation.

So, how are students motivated and how does this motivation differ from that of crowdsourcing participants or practitioners? Are they motivated by an intrinsic desire to do well, extrinsic motivations, such as money, or possibly a negative consequence from lack of effort? Even in a laboratory environment, with an appropriate payment structure, how can researchers ensure participants are cognitively engaged versus absentmindedly cruising through an experiment? The ability to leverage one’s knowledge and intellectual capacity is a key characteristic unique to the knowledge worker, so ensuring cognitive engagement is imperative in studies that are guiding the development of the science behind knowledge work.
In determining the best subjects for experiments, it is important to weigh the potential benefits and challenges posed by each of the following common subjects: students, participants from crowdsourcing sites (such as Mechanical Turk™), and practitioners from the field of study.

For two primary reasons, students are often the most frequently used participants for experiments. First, they are utilized simply out of convenience. It’s generally perceived to be easier to recruit students from the college campuses where the students and researchers reside. Second, from a financial perspective students are generally more practically incentivized with the relatively low financial resources typically available to faculty conducting laboratory experiments. When considering a study focused on the highly cognitive tasks synonymous with knowledge work, while students might be considered a convenience sample for a particular study, they might also be a very reasonable choice as they regularly participate in knowledge-intensive tasks. They are taking in new information, synthesizing it, and generating new knowledge when they complete their coursework, sit for exams, and participate in classroom discussions. In theory, they could be ideal candidates for this type of research. On the other hand, while the use of students has been an often-debated aspect of laboratory experiments, several behavioral studies have shown no statistically significant difference between the performance of students and members of the representative population under study, usually managers (Plott, 1987; Ball & Cech 1996; Bolton et al. 2012; Moritz et al., 2013).

The operations literature has published evidence of intrinsically and extrinsically motivating students in laboratory experiments. The general consensus among research on extrinsic motivation (i.e. – cash, course credit) illustrated that cognitive effort of study
participants increased when given in direct proportion to how well participants perform (Schweitzer & Cachon 2000; Bolton & Katok, 2008; Bostian et al., 2008). Alternatively, other researchers found success with intrinsically motivating cognitive effort (Lepper et al. 1973; Read, 2005). Despite the fact that students regularly participate in knowledge-intensive tasks as they complete their coursework, we struggled with gaining sincere cognitive engagement. In our study, we relied on a nominal extrinsic reward to encourage students to take part in our study and relied on intrinsic motivation to perform the knowledge intensive task. We did not find this successful and suggest working closely with IRB to establish a means for offering a pay-for-performance incentive in light of the literature that suggests the benefits of this reward structure.

Despite the fact that highly cognitive tasks are a standard of life for students, finding an appropriate intrinsic or extrinsic incentive that motivates students to exert sincere cognitive effort is challenging. Even in the most tightly controlled laboratory protocol with the best designed incentive plan, it is impossible to be certain that a participant is giving sufficient cognitive effort and not mindlessly clicking through an experiment. Within the knowledge-work literature, the authors were only able to find a limited number of cases where any attempt to validate the cognitive effort of participants was documented. In a study by Moritz et al. (2013), they conducted three separate studies evaluating the individual differences in cognitive reflection in newsvendor decision-making. Two of the three studies used business professionals as study participants and the other used students. In a description of one of the studies with the business professionals, where they did not use a pay-for-performance incentive system, they explained how they removed subjects from the study when they either completed the task unusually quickly
or their responses illustrated no changes across time periods. They removed n=6 participants out of a total of n=319 that completed the study. While this is an example of the type of analysis we suggest is necessary, they only described this process being conducted with the first of their three studies. There was no explanation as to whether it was conducted with their other two studies or the number of participants that may or may not have been excluded.

In the study detailed in Appendix A, due to the fact that a pay-for-performance incentive was not permitted by IRB, a metric similar to what was used in Mortiz et al. (2013), the overall time to complete a knowledge intensive task was evaluated as a means to check effort. We had to remove n=6 participants out of a total of n=28 that completed the study for lack of sufficient and sincere effort (Figure 11). Thus, we suggest that there needs to be a greater focus on including “effort checks” as part of the experimental design. Best practice would be that rules for excluding participants are determined *a priori* (to prevent arbitrary data scrubbing) and that the researchers begin reporting details on this process and the excluded participants.

**Figure 11: Data Integrity**

![Data Integrity Figure](image.png)
So, how do students compare to participants from crowdsourcing sites such as Amazon’s Mechanical Turk (MTurk)? These sites were initially developed for use by companies seeking to outsource remedial tasks that computers are currently unable to do, such as transcribe books/videos or classify images.

In general, this is how MTurk works: a “Requestor” (employer) posts a job called a “HIT” (Human Intelligence Task) and “Workers” (employees) can then choose to participate in any HITs for which they are qualified to participate. These sites have quickly drawn the appeal of researchers for their potential for providing quick and easy access to online research participants, minimal costs (recruitment and administrative), and a more diverse participant population than normally seen in typical student samples.

As this is a new source of potential participants, research comparing MTurk and other data collection methods has begun to emerge among behavioral researchers. The primary concern for behavioral researchers with “fast and cheap” data is potential quality problems. Some researchers have determined that MTurk can be used to obtain high-quality data inexpensively and rapidly (Paolacci et al., 2010; Buhrmester et al., 2011; Mason & Suri, 2012; Rand, 2012; Chandler et al., 2014) as long as researchers understand its limitations and manage potential risks.

The following are some of the key recommendations for ensuring high-quality data results from behavioral research on Mechanical Turk: 1) track subjects via their unique worker ID to ensure independent responses (Paolacci et al., 2010); 2) recognize that

---

5 Amazon Mechanical Turk allows Requestors to qualify users before they work on their HITs. The qualifications can be anything from a gender or location based requirement to responding to a series of questions or have a specific profile on Mechanical Turk based on their historical performance on HITs.
participants share information and interact in online communities such as mturkforum.com, Reddit, and Facebook (Chandler et al., 2014); 3) consider employing some form of Frederick’s Cognitive Reflection Test (2005) to gauge cognitive effort (Goodman et al., 2013); and 4) include some form of an “attention check” to assess whether participants carefully read instructions (Mason & Suri, 2012; Goodman et al., 2013; Chandler et al., 2014).

But, in a survey of researchers, “worker attentiveness” was listed as the single greatest concern with MTurk (Chandler et al., 2014) and research on whether participants are cognitively engaged is highly inconclusive. According to Chandler et al. (2014), appropriate attention checks may be sufficient since “…it is not clear that inattentiveness necessarily reduces data quality for all phenomena of interest, nor is it clear that people are especially attentive in any other aspect of day-to-day life.” On the other hand, based on the results of their study, Goodman et al. (2013) suggested, “We caution researchers when using MTurk for studies that require participants to pay careful attention to study materials and instructions.” In their research they found that MTurk participants performed significantly worse than students on effort checks and they concluded, “MTurk participants may not be as motivated as student participants to engage in deliberate System 2 cognitive processing.”

Similar to the studies conducted by Goodman et al. (2013), we encountered similar issues with our study conducted using Mechanical Turk (Appendix A). Not only did

---

6 Two common “attention checks” used in these studies: 1) the Instructional Manipulation Check (Oppenhemer et al., 2009) which asks a question at the end of the study (such as “What was this study about?”) to gauge whether participants carefully read the instructions and 2) Reverse Turing test questions somewhere in the midst of the study (such as “Answer yes, if you are reading this question.”).
27.8% of the n=108 participants miss the “attention check” question\(^7\), but, of the n=54 participants who self-rated their effort on the task as \(\geq 95\) (on a scale of 0-100), over 24% of those participants missed the attention check (Figure 12). So, even among the participants who are self-reporting a very sincere cognitive effort, nearly a quarter of them were obviously not sincerely engaged. At this time, we suggest that Mechanical Turk is likely not suitable for research designed to develop a greater understanding of the science behind knowledge-work phenomena.

**Figure 12: Attention Check**

![Participation Attention Check](image)

Finally, how do students compare to practitioners? Students are generally recruited over practitioners due to costs and accessibility. But an often-debated aspect of laboratory experiments is around the similarity between how students and practitioners behave. How well does the work of students translate to the work of practitioners? As noted earlier, several behavioral studies have shown no statistically significant difference between the performance of students and practitioners (Plott, \(\ldots\))

---

\(^7\) Question posed to participants: “If you are paying attention, answer “Strongly Agree” to this question.”
From a cost perspective, practitioners have generally been found to require greater compensation than students since a professional’s time is perceived to be worth more than a student’s, or they are not directly motivated by money (Katok, 2011). In the study conducted by Bolton et al. (2012), while they found no difference in performance of the participants, they paid managers four times more than students. In the study by Moritz et al. (2013), they found that students and managers behaved similarly but they used different incentive methods. Students were motivated with a pay-for-performance structure while managers were not directly compensated based on performance. In our recent study (Appendix A), the most sincere cognitive effort was found when participants from the field of study were used (emergency department physicians). Only n=1 participant was excluded from the set of n=28 total participants for lack of effort (the physician was texting during the study). The physicians in this study were not presented with a pay-for-performance incentive, but given the option to choose between a $5 gift card and a charitable donation to a medical education fund.

While these three studies may suggest that practitioners are motivated more intrinsically and pay-for-performance may not be appropriate, how these types of subjects compare to students in a time when students and practitioners alike are subject to concerns with continuous partial attention is unclear. We suggest that perhaps intrinsically motivated subjects might be better suited for studies involving knowledge-intensive tasks, but more research is needed. Additionally, it is noteworthy that only two of these three studies documented attempts to account for cognitive exertion. When considering research in knowledge work, ensuring sincere cognitive engagement is
critical and we suggest that researchers not only validate participant effort but also document this important information.

4.3 Discussion

The US economy is shifting from a dependence on the productivity of manual laborers to one dependent on the productivity of knowledge workers. Knowledge work is important and there is significant need to further develop the science behind, and remove the barriers to, efficient and high-quality knowledge work. While a relatively limited amount of research on the productivity of knowledge work has been conducted to date, it fits well with the empirical focus in the growing field of behavioral operations management. Many different empirical methods may be appropriate for investigating knowledge-worker productivity, but well-designed experiments that are thoroughly grounded in the literature and have a rigorous approach to data analysis are much needed, are particularly well suited to the task given their strength for precision and control, and have the potential to significantly influence the goal of developing the science behind knowledge-worker productivity.

Ultimately, studying the complex cognitive tasks associated with knowledge work is nontrivial and we discussed the unique challenges introduced by the knowledge workers themselves. Motivating sincere cognitive effort is a key to studies focused on developing the science behind knowledge work. Will extrinsic motivation, as seen in prior studies employing a pay-for-performance incentive, continue to be sufficient to ensure sincere cognitive effort? Or will intrinsic incentives become more important in our changing environment? Through a discussion of the three most common experimental
participants (students, practitioners and crowdsourcing participants) we have made three primary suggestions.

Our first suggestion is for more studies focused on developing our understanding of the right mix of incentives to motivate sufficient cognitive effort. Our second suggestion is for greater effort towards validating and documenting cognitive effort of study participants. Only a limited number of studies have published this information, which leads us to wonder whether it is not being validated or if data scrubbing is actually going on but not being documented. Finally, based on current research, we suggest that the use of crowdsourcing sites may not be appropriate for studies of knowledge-intensive work. But, further studies to test this notion are needed.
4.4 APPENDIX A: Experiment Details

Three laboratory experiments were conducted to evaluate the effects of emphasis framing on knowledge-intensive tasks. All of the details about these studies is not included here, but sufficient information is presented to illustrate the differences in participant effort across the studies.

**Study A** was conducted with n=28 undergraduate business students. Per IRB requirements, students of the authors were not permitted to be recruited and all students were recruited via flyers. The study was administered in a controlled lab-environment where noise and interruptions were minimized. Participants were presented with multiple screens of interrelated information they had to review and use to make a decision at the end. They were presented an “on-line dating” scenario with four key sets of information that was presented over multiple screens: online dating profiles for two hypothetical individuals with information about dining preferences, calendars for the two online daters, information about 10 locations where the online daters could potentially meet, and a local transportation schedule. Subjects analyzed and processed a large quantity of complex, computer-based information to determine the day, meal, and location for two individuals to meet that was optimally aligned with their preferences and minimized their total travel expenses.

**Study B** was conducted with n=108 participants from Mechanical Turk. They were presented the same cognitive task as in Study A, but with the addition of two questions geared toward gaining an understanding of the cognitive effort of participants. First, in
the middle of the information presented, there was a question that stated “If you are paying attention, answer “Strongly Agree” to this question.” Second, at the end of survey they were asked “On a scale from 0-100, how much effort did you put into taking this survey?” In an attempt to gain a sincere response, they were also informed that their responses would not affect their pay for the task.

**Study C** was conducted with n=28 emergency department physicians. They were presented with a hypothetical situation for a patient that arriving via air care. Following the air care message, the physician was instructed to sign into Epic to review the patient’s historical record. (This was an actual patient that had been in this hospital and we obtained IRB approval for the purpose of a chart review.) Upon completion of the chart review, using Qualtrics, the physician documents the patient’s anticipated clinical course and recommended diagnostic and therapeutic interventions, including: top items in the differential diagnosis; anticipated clinical course for the patient in the first hour; and key factors influencing the physician’s recommendations.
4.5 REFERENCES


5

THE OPERATIONAL EFFECTS OF INFORMATION OVERLOAD ON CLINICAL DECISION-MAKING

5.1 Introduction

The healthcare industry is one of the largest and fastest growing industries in the United States. But two landmark reports from the Institute of Medicine (IOM) painted a picture of a healthcare system wrought with quality and efficiency problems and prompted a national focus on healthcare reform. In 1999, “To Err is Human: Building a Safer Health System” exposed the prevalence of medical errors and their implications on the overall cost and quality of healthcare in the United States. In 2001, "Crossing the Quality Chasm" made an urgent call for a redesign of the American healthcare system with a focus on providing high quality (safe, effective, evidence-based, and patient-
centered) care that is delivered in an efficient and equitable manner. A key finding in both reports was the potentially important role of health information technology. The federal government committed an unprecedented $27 billion to promote and expand the adoption of electronic health records (EHR) through the Health Information Technology for Economic and Clinical Health Act (HITECH) of 2009, which has created a 600+% increase in the adoption of EHR systems by hospitals since 2008 (Adler-Milstein et al., 2014), but realized improvements have thus far fallen short of expectations (Black et al., 2011; Jones et al., 2012; Kellermann & Jones, 2013). This apparent contradiction between investment in these operational technologies and lack of realized improvement not only highlights the fact that we do not adequately understand the factors that contribute to efficient and effective work systems, but, most alarmingly, the quality of patient care may also be suffering.

A 2013 study found that over one-third of physicians reported missing test results in an EHR system because they are simply overwhelmed by data and information (Singh et al., 2013). That study highlighted a significant problem facing physicians today: information overload. While the most-informed decision is often the best decision, the theory of bounded rationality argues that humans have only a limited capacity to process complex problems and information (Simon, 1957). Up to a certain point, decision-making performance is positively correlated with the amount of information a decision-maker receives. But, beyond that point, the information-processing requirements of a task exceed the information-processing capacities of the decision-maker, sending him/her into a state of information overload (Eppler et al., 2004). As a result, decision-making performance decreases. In healthcare, EHRs provide data and
information that could potentially increase the quality and efficiency of clinical decision-making and improve patient care. But, the evidence suggests that EHRs are currently failing to meet those objectives and, even worse, potentially contributing to errors in clinical decision-making (Ash et al., 2004; Kuperman, 2011; Singh et al., 2013). The role of information overload on the timeliness and quality of clinical decision-making performance needs to be better understood.

In this study, we introduce and test “emphasis framing” as an operational tactic to help mitigate the effects of information overload on the quality and timeliness of clinical decision-making. Emphasis framing occurs when some aspect or component of the information being exchanged is highlighted or stressed to make it more easily, or likely to be, processed by the recipient (Entman, 1993; Druckman, 2001). From a cognitive perspective, a “frame” serves as a simplifying structure for cognitive categorization (Davies and Mabin, 2001), which decreases cognitive load and potentially improves decision-making performance. This study evaluates the effects of emphasis framing as an operational tactic to improve cognitive categorization, decrease cognitive load, and improve the quality and efficiency (timeliness) of decision-making.

5.2 Relevant Literature and Hypotheses

Clinical decision-making is a complex, knowledge-intensive process that involves a careful analysis of harms and benefits associated with different treatment options. These decisions, often associated with high stakes and important long-term consequences, are frequently made in the face of competing priorities, limited resources and information, and an incomplete clinical picture. Physicians are regularly faced with
executing high-quality clinical decision making under these challenging circumstances and the prevalence of medical errors has become well known.

In an attempt to provide data and information that could potentially improve the quality and efficiency of clinical decision-making, and ultimately patient care, a significant investment in health information technology was made in the U.S. But, Jones et al. (2012) contend that, “swapping out of the medical record cabinet and prescription pad for a computer is proving insufficient to realize the benefits of health IT.” According to research from IBM, “every day we create 2.5 quintillion bytes of data – so much that 90% of the data in the world has been created in the last two years alone.” Similarly, the greater than 600% increase in the adoption of EHR systems by hospitals since 2008 has increased the intensity of information being presented to physicians, potentially inducing information overload (Ash et al., 2004; Harrison et al., 2007; Kuperman, 2011; Singh et al., 2013). They are facing a significant problem: information overload.

Information overload occurs when the information intensity increases to a point where the information-processing requirements of a task exceed the information-processing capacities of the individual (Eppler et al., 2004). It is largely determined by the quantity and complexity of information needing to be processed and plays an important role in the ability of a decision-maker to accurately and efficiently process information (Hiltz & Turoff, 1985; Keller & Staelin, 1987; Schneider, 1987; Schick et al., 1990; Speier et al., 1999; Eppler & Mengis, 2004). While the most-informed decision is often the best decision, the theory of bounded rationality argues that humans have only a limited capacity to process complex problems and information (Simon, 1957). Up to a certain point, decision-making performance is positively correlated with the amount of
information a decision-maker receives. But, beyond that point, the information-processing requirements of a task exceed the information-processing capacities of the decision-maker, sending him/her into a state of information overload (Eppler et al., 2004). As a result, decision-making performance decreases.

This concept of information overload is readily seen in clinical decision-making. According to Croskerry et al. (2013), it is the “flaws in clinical reasoning rather than lack of knowledge that underlie cognitive errors”. Drawing from Kahneman’s dual-process theory of decision making, they conclude that the majority of clinical errors occur “on the front line” when clinicians are under immense time pressures and resources are in short supply. Under those trying conditions, when the information-processing requirements of a task exceed the information-processing capacities of the decision-maker, they posit that physicians are subconsciously looking for shortcuts or ways to reduce the excessive cognitive load (information overload) and they revert to trained, procedural knowledge executed with System 1 thinking, instead of the System 2 mode of thinking that is more appropriate for the complex, knowledge-intensive nature of clinical decision making.

For example, a physician’s perception of patient risk may be influenced by whether their outcome is expressed in terms of the probability that the patient might live or die (e.g. – “There is a 40% chance a patient will survive” vs. “There is a 60% chance a patient will die”). If that situation is presented to a physician that is cognitively overloaded, they may be influenced by how a problem is presented in one of a variety of logically equivalent alternatives, or by an equivalency frame, and there is an increased likelihood of errors in clinical decision-making (Croskerry, 2003). Ultimately, in order to
reduce clinical errors and improve decision making, we need to find ways to reduce
cognitive load. This will better enable physicians to avoid the pitfalls associated with
using intuitive System 1 thinking and move more deliberately into the System 2 mode of
thinking that is more appropriate for the complex, knowledge-intensive nature of clinical
decision making.

According to the medical literature, one strategy for achieving this is to increase the
speed and reliability of feedback to decision-makers. They suggest that, since effective
clinical decision making requires convergence on the higher level abstractions of
information and negotiations in order to develop a shared understanding of information,
enhanced collaborative thinking and the assistance of decision-support tools to
augment an individual physicians own knowledge and capabilities are necessary
(Croskerry, 2003; Graber et al., 2012).

Unfortunately, research has shown that instead of helping solve this problem, EHRs
may actually be contributing to the problem and potentially even decreasing the quality
of clinical decisions. EHRs have demonstrated the ability to effectively convey
information about patients and their medical history and capture procedures to generate
better charges. But, in direct contrast to the need for enhanced collaborative thinking
and the assistance of decision support tools to augment an individual physicians own
knowledge and capabilities, research from the Institute of Medicine (2012) suggest
EHRs have actually reduced the type and amount of direct interactions between health
care providers. This apparent contradiction highlights one of the many reasons we may
be seeing a lack of realized improvements despite the significant investment made in
health information technology. But there is hope. Research suggests EHRs have
potential to reduce information overload and significantly improve clinical decision making by better supporting collaborative thinking and promoting feedback (Hamm & Zubialde, 1995; Schiff & Bates, 2010).

To develop an understanding of the role of information systems and the factors that induce information overload and influence decision-making performance, we refer to the information systems literature. First, the characteristics of the task have been shown to influence information overload. Eppler and Mengis (2004) found that a high volume of information, task complexity, and novelty of information increase the information processing requirements of a decision-making task. Thus, it should be no surprise that the complex, knowledge-intensive tasks often associated with clinical decision making can contribute to information overload.

Second, on a more positive note, the characteristics of the decision-maker have been shown to influence information overload (Speier et al., 1999; Eppler & Mengis, 2004). The more qualified and experienced a decision-maker, the faster and more efficient he/she is at processing information. Thus, a more experienced physician is likely to be less susceptible to information overload than a first-year resident.

Finally, the information technology or medium has been shown to influence information overload (Eppler & Mengis, 2004). According to Media Synchronicity Theory (MST), matching the capabilities of a medium, that best supports the conveyance or convergence process, can affect the cognitive load of a decision-maker and, ultimately, the performance of the process (Dennis et al, 2008). For the portions of a decision process where conveyance of large amounts of raw information is required, “individuals will have less of a need to transmit and process information at the same time” (Robert &
Dennis, 2005) and media that supports low synchronicity is most appropriate. The logic behind this media choice lies in the fact that the conveyance process allows for individual consumption and processing of new or diverse information with decreased interaction between individuals. As the research from IOM suggested, the fact that EHRs have reduced the type and amount of direct interactions between health care providers, suggests EHRs are primarily a media with low synchronicity that is best suited for conveyance of information. But, media with high synchronicity is more appropriate for convergence as it generally requires fewer cognitive resources than conveyance and ultimately may reduce information overload (Dennis et al., 2008). Thus, perhaps this provides further evidence that the use of an EHR system as a means to primarily convey information to health care practitioners will likely never contribute to improved clinical decision making and may even make it worse. Research has shown that the appropriate use of synchronous vs. asynchronous media, as a means of information exchange, can influence cognitive load (Dennis et al., 2008). And using EHRs to better match the appropriate medium to the conveyance or convergence process may reduce information processing and convergence time while also improve decision-making performance.

In this study, we test the effects of “emphasis framing” on the quality and timeliness of clinical decision-making. Emphasis framing occurs when focus is placed on some aspect or component of the information being exchanged to encourage certain interpretations or provide meaning and appropriate context (Entman, 1993; Druckman, 2001). Emphasis frames also help giving meaning to information being exchanged and promoted shared understanding and convergence on ideas. From a cognitive
perspective, a “frame” serves as a simplifying structure for cognitive categorization (Davies and Mabin, 2001), which decreases cognitive load and potentially improves decision-making performance. Therefore, this research hypothesizes:

\[ H1: \text{In knowledge-intensive work environments, emphasis framing reduces decision-making time.} \]

\[ H2: \text{In knowledge-intensive work environments, emphasis framing improves decision-making quality.} \]

5.3 Methods

In this study, we conducted a controlled laboratory experiment designed as a balanced, fixed-effects means model. We measured the effect of emphasis framing on two operational performance metrics when physicians are under information overload: (1) the quality (accuracy) of the physician’s clinical evaluation, and (2) the efficiency (timeliness) of his/her clinical decision-making.

5.3.1 Sample Design and Selection

Because our study seeks to measure different metrics related to clinical decision-making, we identified physicians as our target population. In an effort to execute this study using a real EHR system, it was determined that all participants would need to have access to not only the same EHR system but equal access to the same set of historical charts. In order to increase the reliability of our results, we limited our sample to one medical specialty. While all board-certified physicians within a hospital share some level of common knowledge, the practice of medicine within each specialty varies
significantly. This sampling design enabled us to reduce the likelihood of introducing random measurement error as all participants would be familiar with the terminology and best practices upon which clinical decision-making would be evaluated. We designed our study for emergency medicine physicians due to the fact that they primarily focus on immediate decision-making and action in response to acute illness and injury. Our last consideration in our sample design was to take into consideration the difference in physician skill level. Therefore, we used a randomized stratification to assign participants into two treatment groups by skill level (resident vs. attending).

A total of 28 emergency department physicians from an urban academic emergency department agreed to participate in the study. This sample represents 28% of the eligible participants from this emergency department. Of the 28 individuals who participated in the study, n=4 were excluded for the following reasons: two physicians were forced to quit mid-study due to technical issues that prevented them from accessing their EHR system; one physician was interrupted mid-study to address an unexpected work related issue; and one physician encountered a technical issue with the software tool used to conduct the experiment. Of the remaining 24 physician participants, they were randomly stratified by treatment group (12 per group) by skill level with 9 attendings and 3 residents in each treatment group. The average experience across all participants was 8.9 years (SD = 6.2) since medical school graduation. In our physician sample, 29.2% were female and 70.8% male. In the entire population of emergency physicians, 23.50% are female and 76.5% are male (census data). Due to our sample size, we were unable to test if there was a statistically
significant difference by gender (our data failed to meet the underlying condition that assumes the distribution of the sample is approximately normal \(i.e. np*(1-p) \geq 100\)).

The subjects were recruited via email sent out by the department’s medical director, to all physicians in the department of emergency medicine. In an attempt to mitigate the effects of response bias, they were invited to take part in an EPIC\textsuperscript{8} usability study. We felt that notifying participants that we were examining the effects of information overload might lead a participant to behave in a manner different than their true response to the amount of information being presented. Of specific concern was the potential for social desirability bias. Since we were measuring the quality and timeliness of their decision-making, we were concerned participants would unnaturally adjust the amount of time spent reviewing the information presented in an attempt to mitigate the effects of “information overload”. As a recruitment incentive, subjects were offered the choice of a $5 Starbucks gift card or a $10 contribution to a medical education fund. The amount selected for the incentive was chosen as a value that is high enough to pique interest and be viewed as a “thank you” for their time without creating any survey response bias.

### 5.3.2 Experimental Protocol & Design

This is a detailed study conducted in an environment not familiar to many individuals outside the medical community, and perhaps even outside of the emergency department. In order to best explain this experiment, we will first provide a high-level

---

\textsuperscript{8} EPIC is a privately held healthcare software company. The urban emergency department where this study was conducted uses its EHR software.
overview of the experimental protocol. This will then be followed by a detailed description of our experimental design.

5.3.2.1 Overview of the Experimental Protocol

This study was individually administered in a controlled lab environment where distractions and interruptions were minimized (Image 1). Upon entering the laboratory environment, the principal investigator provided an overview of what the participant would be doing during the study and informed the participant that the PI would not be staying in the laboratory during the experiment, but would be readily available just outside the laboratory, in order to assist if necessary. The participant was provided with a pen and paper in the event he/she wanted to take notes.

Physicians began the experiment by clicking the “start” button on the computer screen. The first screen to appear provided the IRB-required information sheet and consent. In the interest of study integrity, the second screen requested confirmation that the participant would refrain from discussing any aspect of the experiment with others. The third screen collected demographic information. The fourth screen provided the participant more explicit details as to what he/she would be doing during the experiment and introduced the details of the problem (Appendix B1). On the fifth screen, the physicians clicked on a link to listen to a pre-recorded “air care” report, from one of the department’s Air Care physicians, describing the hypothetical patient that would be arriving via helicopter in a few moments (Appendix B2). Following the message, on the sixth screen, the physicians were provided the medical record number for the arriving patient and instructed to use their regular credentials to sign into EPIC and conduct a
For their final step in the experiment, the physician documented his/her expectations for the patient’s clinical course and his/her recommended diagnostic and therapeutic interventions (Appendix B5).

**Image 1: Physician participating in experiment**

![Image of a physician using a laptop](image)

5.3.2.2 Experimental Interface

Two experimental interfaces were used to conduct this study. The first interface was provided by Qualtrics (Provo, Utah: v. 2009) survey software. This software was first tested in initial pilot studies. For this experiment it was used to guide the participants through the multiple phases of the study, to collect data on how much time each physician spent on each aspect of the study, and a place for physicians to document their clinical decisions.

The second experimental interface was provided by EPIC, one of the most commonly used EHR systems in the United States (Off. of the Nat. Coord. for HIT, 2015). We received IRB approval for the physicians to conduct a retrospective chart
review (RCR), also known as a medical record review, for a specific patient. For patient safety, the medical record was viewed by physicians in a “read-only” environment that does not allow for any modifications to be made to the medical record.

5.3.2.3 Study Design & Development of Experimental Instrument

In this study, we conducted a controlled laboratory experiment designed as a balanced, fixed-effects means model. Our control group was presented with all information needed to make a decision, but it did not receive an emphasis frame. The treatment group was presented with all of the same information as the control group, but was provided with additional information that provided an emphasis frame. The emphasis frame highlighted an important piece of information found in the EHR, but which could be easily overlooked due to the amount of information present in the patient’s chart. We measured the effect of emphasis framing on two operational performance metrics when physicians are under information overload: (1) the quality (accuracy) of the physician’s clinical evaluation, and (2) the efficiency (timeliness) of his/her clinical decision-making.

In order to develop a valid and reliable experimental instrument, we worked in collaboration with the following three individuals employed by the emergency department where the study was conducted: a 4th year resident who was an air care lead; an attending physician who has a master’s degree in clinical education and also serves as an education and training instructor within the department; the medical director; and the vice chair of research for the department of emergency medicine (also a co-author of this study). Each of these individuals played an integral role in our
development of the experimental instrument, they were essential not only to the
development of the clinical aspects of our study, but also for their education and training
instruction within the department.

Emergency medicine physicians regularly participate in educational training, such
as simulation training for practicing clinical scenarios; journal clubs to discuss new and
relevant clinical literature in their field; and morbidity and mortality (M&M) conferences
consisting of non-punitive, peer reviews of past events that occurred during the care of
patients. Thus, it was important to develop our instrument with an understanding of the
situations in which these physicians are used to practicing and/or discussing medical
decision-making, but also to ensure that we did not introduce response bias by selecting
a clinical scenario that had been a specific focus of a previous clinical education training
within this department.

Each aspect of our study, described below, was designed to ensure the reliability
and validity of the experiment. We developed a hypothetical patient profile and clinical
presentation to the ED based on the knowledge that the physician participants in our
study would employ a Bayesian approach to medical decision-making (i.e., considers
the values and costs associated with potential outcomes). The hypothetical patient was
involved in a minor motor vehicle crash (MVC) while he was restrained in his safety belt
and did not exhibit any obvious extremity trauma. This information was important to
highlight a low probability of severe injury. The patient was fairly young, with a history of
alcohol related injuries (as noted in his medical record), but an otherwise unremarkable
medical history (e.g. - no history of diabetes, cardiovascular disease, seizures or
neurological issues). Thus, a low probability for significant complications from these
diseases. The patient was noted to be agitated on scene, which was most likely due to alcohol intoxication (because of patient history) or information presented in the initial physical exam could also indicate the possibility of traumatic brain injury and/or intracranial hemorrhage. Noting that the patient was agitated on scene was important because the Air Care team would need to intubate him for airway protection. Succinylcholine (a.k.a. “sux”) is the depolarizing paralytic agent used by this Air Care.

While there will always be some uncertainty about the most appropriate clinical strategy, up to this point, all details of our experiment were designed such that the hypothetical patient presentation, history, and physical exam are fairly straightforward and unremarkable. Thus, the top items in a physician's differential diagnosis and the anticipated clinical course for the first hour upon arrival would be fairly consistent across all study participants.

However, one key detail significantly increases this patient’s morbidity and mortality risk. Unknown to the Air Care physicians, the patient has a genetic condition called malignant hyperthermia (MH), which creates a potentially dangerous reaction to succinylcholine. It is noted that the patient begins to exhibit tachycardia, an abnormally rapid heartbeat, and signs of muscle rigidity in his extremities after intubation. For physicians in the treatment group, the participants are informed that the patient’s wife mentioned that he has a history of bad reactions to anesthesia, which is designed to serve as the emphasis frame. Highlighting this information (which is located in his medical record) should help the physician more quickly and accurately identify that the rigidity and tachycardia are most likely attributed to succinylcholine-induced malignant hyperthermia and not any of the significantly less risky complications related to the
MVC, such as shock, seizure, or neuroleptic malignant syndrome (NMS). This information is important because the immediate care plan for an MH patient that has been administered succinylcholine (i.e., timely administration of Dantrolene) is notably different than for a non-MH patient experiencing shock, seizure, or NMS. If not immediately treated, the body fails to supply sufficient oxygen to the body, remove sufficient carbon dioxide from the body or appropriately regulate their body temperature, which ultimately can lead to circulatory collapse and death.

The physician participants were presented the clinical details for this experiment in order to inform their clinical decision-making. First, they were presented a pre-recorded message from one of their air care physicians. The script of that message was as follows:

“Air Care 1 will be en route with a 43-year-old male restrained driver in a motor vehicle accident who is agitated on scene and was unable to be controlled and intubated for airway protection. He is notable to have a heart rate of 130 and blood pressure 150 over 87. A respiratory heart rate of 28, satting 99%, his breath sounds are clear bilaterally, his belly is soft, he has no obvious extremity trauma. He is noted to be somewhat rigid after the intubation. Of note, his wife did mention that he has a history of allergies to anesthesia. ETA is approximately 5 minutes. Any questions?”

Second, after they heard the air care message, they reviewed the patient’s historical medical record in their EHR system.

---

9 The italicized sentence is the “emphasis” frame and was only presented to physicians in the treatment group.
There were several reasons we selected to present the situation to the physician participants via pre-recorded Air Care message. First, an emergency physician’s perception of a patient’s clinical condition tends to vary depending on the source of the information, especially with out-of-hospital intubations. Research from the emergency medicine literature has shown that out-of-hospital intubation by paramedics not only does not improve morbidity or mortality, but also may lead to unfavorable clinical effects, adverse events, and errors (Wang & Yealy, 2006). Thus, to increase the likelihood of consistent interpretation about the hypothetical patient’s clinical condition, it was important that the patient arrive via Air Care, which guarantees the patient’s out-of-hospital care would have been provided by an emergency department physician and not a paramedic.

Additionally, to further increase the likelihood of consistent interpretation, it was important to make it clear that the message was delivered by one of their own physicians. The message was recorded by the physician that most frequently makes that call in their actual ED environment. Selection of this specific physician was important for two reasons. First, the physicians do not usually individually identify themselves when making these calls in the actual ED, simply which helicopter they should be expecting (e.g. “Air Care 1”). Thus, we selected the voice they would most likely immediately recognize and put them in a similar mind set. Second, while information coming from air care usually contains a fairly consistent set of information, there is not a specific script they follow. Thus, having the message recorded by the physician that most frequently delivers these message increases the likelihood that it
includes all of the details a study participant would expect to receive about a patient arriving via air care.

Second, when a patient arrives to this emergency department via Air Care, the emergency physicians at the hospital are first notified of about the details of the arriving patient via a phone call from the Air Care physician. While we realize that a phone call would most closely mimic the real ED environment, we chose to deliver the patient scenario via a pre-recorded voice message for two primary reasons. First, it helps to increase the external validity of our experiment as it closely resembles how they gather this information in the real environment, but also enables us to ensure the information presented to participants is consistent throughout the experiment. We know that the exact same information is delivered in exactly the same way (speed of voice, intonation, etc...). Second, it would be extremely difficult, from a logistical perspective, to find a way to schedule the specific physician needed for this part of the study, in the experiment location, at times when all of the study participants were available.

Following the Air Care message, the physician is instructed to sign into EPIC to review the patient’s historical record. We made the decision to present the historical clinical information in the format they use in their daily practice of emergency medicine, the EPIC EHR system. Not only did this enhance external validity, but it also contributed to the reliability of our experiment because we were confident that the study participants were all familiar with the EHR software for the purpose of reviewing a patient’s medical history. While the physician participants all had the required credentials to access patient information, we were required to acquire specific IRB approval to repeatedly access an individual historical medical record for the purposes of a research study.
Once we received IRB approval, we were able to request a list of any historical cases that met the criteria established in our experimental protocol, have been to this emergency department since EPIC implementation, and were still alive today. Once we received the list of potential cases, two physicians independently reviewed the potential candidates and the most appropriate record was selected for use with this study\textsuperscript{10}.

It is important to note that when emergency physicians are working clinically in the emergency department, due to the high volume of patients they see, it is more common that they would conduct their own physical exam on a patient prior to reviewing their medical record. While it might have been feasible to use a medical simulation (i.e., “dummy”) to replicate this scenario in our experiment, it was determined this form of information gathering would introduce of a significant amount of unnecessary variability.

Upon completion of the chart review, using Qualtrics, physician were informed to “Assume no changes in the patient’s clinical exam” and to document the patient’s anticipated clinical course and recommended diagnostic and therapeutic interventions, including: (a) top items in the differential diagnosis; (b) anticipated clinical course for the patient in the first hour; and (c) key factors influencing the physician’s recommendations. These three open-ended questions were written in conjunction with the emergency physician educator based on standard terminology used to capture clinical decision-making. The request to “assume no changes in the patient’s clinical exam” were important for instrument reliability. In the practice of emergency medicine, it is possible for a patient’s clinical condition to rapidly change in a short period of time. Thus we wanted to ensure the physician participants viewed the presented information

\textsuperscript{10} We received a waiver of authorization since the focus of this research was on medical decision-making and not the patient and, therefore, posed minimal risk to the privacy of the subject.
as a “snap shot” to avoid any effects from a time series bias in their clinical decision-making.

Prior to the start of our data collection, we worked with two emergency department physicians to conduct a pilot test and adjusted our experiment based on two key findings. First, the “button” that physicians clicked to play the air care message was too small and led to confusion as to what they were supposed to do on that page of the experiment. And second, while it was noted in two different places that the participants were to review the patient’s “historical” record, both noted that it was important to note that no information about the hypothetical situation will be in the medical record (i.e., it should be treated more as a retrospective chart review).

This study was individually administered in a controlled lab environment where distractions and interruptions were minimized (Image 1). The first author was present for all data collection activities and followed a specific and consistent protocol with each participant to reduce the likelihood of variations in administration. The PI was available in the event any issues arose during the study. There were no significant events that occurred during or after the experiment was underway.

5.3.2.4 Measures

In our study, we measured the effect of emphasis framing on two operational performance metrics when physicians are under information overload: (1) the quality (accuracy) of the physician’s clinical evaluation, and (2) the efficiency (timeliness) of his/her clinical decision-making. To evaluate the timeliness of decision-making, the Qualtrics survey tool provided information on how long a physician spent during each
phase of the experiment. We collected data on all aspects of the experiment but were specifically interested in how much time each physician spent on the following activities: problem description, chart review, and decision-making. We defined decision-making time as the duration from when the physician completed the chart review to the time when documentation of clinical decision-making is completed.

For methodological rigor, based on a pre-established quality rubric, four independent physicians independently scored the quality of the physicians’ decision-making. The scoring rubric is based on best practice as determined by emergency medicine physicians familiar with the scenario and patient history being presented in this study, and reflects completeness, accuracy, recognition of the underlying problem, and identification of the correct protocol moving forward (Appendix B6). The four quality-related questions were as follows: A) Quality of the physician response to possible causes for the patient’s altered mental status/agitation on scene; B) Quality of the physician response in terms of addressing the rigidity of the patient post intubation; C) Overall quality of the physician decision-making; and D) Overall clarity and thoroughness of the physician response. The scorers do not know each other, were not affiliated with the experiment, and independently conducted their evaluations (Image 2). They were instructed to score each response, on a scale from 0-100, for each of the four questions. To increase inter-rater reliability and to reduce experimenter bias, they were presented with clearly stated scoring guidelines for each question. After collecting all of their responses, the individual experimental participant’s scores to the four questions were weighted A = 25%, B = 40%, C = 25%, and D = 10%. These weightings
were selected based on their potential contribution to the morbidity and mortality (outcomes) for this hypothetical patient.

**Image 2: Physician scorer**

---

5.4 Results

5.4.1 Quality of Clinical Decision-Making

Consistent with our hypothesis, emphasis framing increased the quality of clinical decision-making. We started our analysis by measuring the inter-rater reliability (IRR). For each of the four scores, we chose to calculate the intra-class correlation coefficient (ICC), which measures the proportion of variance of an observation due to between-subject variability in the true scores (Ebel, 1951). The higher the ICC values, the greater IRR, with an ICC estimate of 1 indicating perfect agreement and 0 indicating only random agreement. It is the most appropriate measure of inter-rater reliability for interval scale data. With 99% confidence, our results indicate high inter-rater reliability across all four questions with an average of 86.9% variance explained (Table 14).
Table 14: Intra-Class Correlation Coefficient

<table>
<thead>
<tr>
<th>Question</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.832</td>
<td>0.877</td>
<td>0.880</td>
<td>0.886</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The mean score for the control group was 44.6 (SD 24.1) and 70.2 (SD 30.0) for the treatment group (Table 15). According to Levene’s test for homogeneity of variance (p>=0.05), it was determined that the variances did not differ significantly (Table 16). Thus, we conducted an analysis of variance (ANOVA) and determined there was significant main effect for the treatment; F(1, 24) = 5.313, p = 0.03.

Table 15: Quality Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>44.6</td>
<td>24.1</td>
<td>12</td>
</tr>
<tr>
<td>Frame</td>
<td>70.2</td>
<td>30.0</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>57.4</td>
<td>29.7</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 16: Levene's Test of Equality of Error Variances

<table>
<thead>
<tr>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.437</td>
<td>1</td>
<td>22</td>
<td>0.515</td>
</tr>
</tbody>
</table>

We extended our ANOVA analysis to see what effect the emphasis frame has after the effect of our covariates – skill level, defined by the number of years since medical school graduation, and gender. We were not able to able to attain statistical significance when conducting our analysis of covariance (ANCOVA) with gender as our co-variate. With 90% confidence we were able to determine there was a significant effect of emphasis framing after controlling for the effect of skill level, F(2,24) = 2.605, p = 0.098.

Finally, we conducted a Chi-Square test to determine if the percentage of participants that correctly identified the hypothetical patient as potentially having MH
differed by whether they had the emphasis frame or not. We found that only 4 participants in the control group correctly identified the MH, but 9 participants in the treatment group. We found that the number of participants that correctly identified the MH patients did differ depending on whether an emphasis frame was provided, $X^2 (1, N=24) = 4.20, \ p = 0.04$.

5.4.2 Timeliness of Clinical Decision-Making

Contrary to what we expected, the timeliness of clinical decision making, defined as the duration from when the physician completed the chart review to the time when documentation of clinical decision-making is completed, increased with the emphasis frame. The mean time for the control group was 266 seconds (SD 69.1) and 421 sec (SD 24.8) for the treatment group (Table 17). According to Levene’s test for homogeneity of variance ($p<=0.05$), it was determined that the variances are significantly different (Table 18). Thus, we conducted a two-sample t-test assuming unequal variances and concluded that emphasis framing increased the average clinical decision-making time, $t(12) = -1.86, \ p < 0.05$.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>265.6</td>
<td>69.1</td>
<td>12</td>
</tr>
<tr>
<td>Frame</td>
<td>420.7</td>
<td>245.2</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>343.2</td>
<td>193.5</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 17: Timeliness Descriptive Statistics

Table 18: Levene's Test of Equality of Error Variances

<table>
<thead>
<tr>
<th></th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>8.321</td>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>
5.5 Conclusions and Limitations

This research focused on the cognitive limits and, specifically, the influence of information overload on clinical decision-making by emergency medicine physicians. The results of this research provide us with some new insight as to how emphasis framing as an operational tactic might help better facilitate high-quality clinical decision-making when physicians are experiencing information overload. Current design and utilization of EHRs are primarily used for the conveyance of information. Ideally, healthcare information technology systems should not only collect information, but better support clinical decision-making. We posit that the process a physician goes through in order to ultimately determine the best plan of care for a patient requires a significant amount of convergence, which focuses on the higher level abstractions of information and a perhaps a more synchronous media may be more appropriate.

Ensuring timely access to medical care is an important goal for any healthcare setting, and especially in the emergency department where physicians are often providing acute care. The Joint Commission, the largest accreditation and certification organization for health care organizations in the U.S., has highlighted the importance of timeliness of care as a key component of several of their Core Measures. As an example, for patients arriving to the emergency department with an acute myocardial infarction (acute MI), a patient should receive a percutaneous coronary intervention (i.e., angioplasty with stent) within 90 minutes of hospital arrival.

We hypothesized that an emphasis frame would enable physicians to navigate the EHR more quickly and ultimately provide more timely care. Surprisingly, in our experiment, the emphasis frame was found to increase clinical decision-making time.
But, our results also found that an emphasis frame increased the quality of clinical decision-making. These results suggest that perhaps there is a conflict in our objectives and there is actually a quality and timeliness tradeoff such that faster decision-making actually impedes the careful consideration necessary for high-quality decision making. This result raised some interesting questions that would benefit from further research to further understand the potential trade-off between timeliness and quality.

Our results did suggest that emphasis framing has the potential to improve clinical decision-making. However, these findings should be considered in light of several limitations. Our results were based on data from a single, urban, academic emergency medicine department using EPIC EHR. It is not clear whether our results would hold in different medical specialties or even different EHR systems. Also, the design of our medical scenario was uniquely created to test the effects of interest. The limits of emphasis framing are unclear at this time.
CONCLUSIONS, DISCUSSION, AND FUTURE RESEARCH

This results of this research provide us with deeper insight into some of the challenges associated with conducting experiments in knowledge-work and the potential for emphasis framing as a reasonable operational tactic to better facilitate high-quality clinical decision-making for physicians under information overload.

In our pilot studies, with students and crowdsourcing participants, while our results were interesting, due to some challenges, we were unable to attain statistically significant results. Between the difficulties we experienced with IRB, to the challenges with attaining sincere cognitive effort on behalf of the participants, it is clear that studying the complex cognitive tasks synonymous with knowledge work is difficult. We found that subject recruitment is critical when conducting experiments to develop the science behind knowledge-worker productivity.

Our results from the emergency department study showed that emphasis framing improves clinical decision-making quality by aiding physicians in identifying the most
relevant pieces of information, thereby mitigating the effects of information overload. Surprisingly we found that emphasis framing increased clinical decision-making time. While we had hypothesized that it would enable the participant to navigate the EHR more quickly, it appears that perhaps there is actually a quality and timeliness tradeoff such that faster decision-making actually impedes the careful consideration given to high-quality decision making. Our results suggest further research is necessary to better understand the relationship between timeliness and quality in medical decision-making.

From a theoretical perspective, this research contributes to the development of our understanding of decision-making in knowledge-intensive work environments, specifically clinical decision-making. It builds on the OM literature and the cognitive sciences to explore the decision-making process through a behavioral lens in knowledge-intensive work environments. It draws from the operations, information systems, cognitive science, and medical literature to develop our understanding around the role of information overload and decision-making.

There are four principal benefits of this research. First, the research contributes to the development of the science of managing knowledge-intensive work environments. Despite its vast importance to the economy, the operations management literature has largely focused research elsewhere, such as on “the physical tasks in manufacturing, construction, and other industries” (Hopp et al., 2009). But the results of our pilot experiments illustrated, studying the complex cognitive tasks associated with knowledge work is nontrivial and the unique challenges introduced by the knowledge workers themselves needs to be further explored. Second, this research contributes to
the relatively limited empirical research on the influence of interruptions and information overload in decision-making performance. This methodology is helpful for understanding the many behavioral and operational aspects of how people interact with work systems, which has not been widely used in operations research. Third, as far as the author is aware, this is the first test and evaluation of emphasis framing as an operational tactic for mitigating the effects of information-overload on decision making. Finally, this research illustrated some of the challenges with conducting behavioral, knowledge-intensive experiments on crowdsourcing sites and further research is needed to determine if that platform is appropriate for these types of experiments.

As discussed throughout this dissertation, there are several opportunities for extensions of this research. Considering the challenges we faced in our pilot experiments, a need exists for more research focused on understand the best mix of incentives needed to gain sincere cognitive effort from participants in studies exploring the complex cognitive tasks associated with knowledge work. Additionally, further research is needed to develop a better understanding of the unique challenges introduced by crowd sourcing platforms such as Mechanical Turk. From our experiment with the emergency department, our results illustrated a documented a statistically significant benefit from the introduction of an emphasis frame, but vast opportunities exist for extending this research to potential implementations for the improved practice of medicine.
The challenges of dating in 2014!

Adeline and Declan are two young professionals who met via an online dating site. They have viewed each other’s profiles and have decided to set up a first date.

They want to find a day and time, within the next 7 days, to meet up for a meal. But, finding a day and location that fits their busy schedules is no small challenge with two young professionals focused on their careers!

Adeline and Declan live in a country called Zen. They do not have cars, but Zen’s 12 cities are all connected via train.

They can choose to meet in any city for breakfast, lunch or dinner.

The meeting day, time, and location needs to not only accommodate their busy schedules, but should be aligned with their preferences from their online dating profiles (shown next).

Additionally, since they are both cost-conscious business professionals, they want to meet in a location that minimizes their total travel expenses (combined cost of their rail tickets).
A2: Dating Profiles

Adeline's Dating Profile

Basic Info
Name: Adeline
Age: 26 Years Old
Location: The City of Alpha
Profession: Actuary

About Me
I'm fun, energetic, outgoing, analytical, independent, and have a good sense of humor. My interests include international travel, fitness, spending time with my family, any sort of outdoor activity (hiking, camping, etc...), science, and seeking adventure. My favorite foods are Mexican and Thai. I've traveled to 12 countries on 4 continents. I once memorized Pi to 100 digits, for literally no reason at all. My favorite place in the world is La Fortuna Waterfall in Costa Rica. I don't believe in luck, you create your own destiny and choose your own happiness. Albert Einstein is my hero and I keep an action figure of him on my desk at work. I'm a night owl by nature and love mornings only when they are at the tail end of an adventure-filled evening. I think I was born with an overactive sarcastic gene... and in case you were wondering, no that gene does not actually exist, but I think it should. If you think so too, send me a message!

Declan's Dating Profile

Basic Info
Name: Declan
Age: 28 Years Old
Location: The City of Omega
Profession: Manager of Consumer and Market Knowledge

About Me
Welcome to my world! I'd like to think it's a pretty good place to live. My ideal day would be spent hiking through the mountains, rafting in the Adobe River, cheering on our beloved Silverbacks to victory, cooking an authentic Italian meal from recipes handed down to me from my Sicilian Grandmother, or whatever other cool activity we chose to partake in. You see, I'm not looking for someone to join me on my journey through life that isn't up for a bit of a challenge. I don't care much for sitting around the house, watching TV. To be honest, I don't even know why I own a TV. I suppose it looks kinda cool next to my fireplace. I'd rather be out experiencing life here in Zen or out on the road checking out the World's Largest Ball of Twine or tasting our way through the small, family owned restaurants, brewerias, and wineries spread throughout our great country - as long as we leave the Italian food to a meal cooked at home, nothing compares Grandma's recipes that have been masterfully crafted over the years :) While my weekends are full of physical adventures, I thrive on the analytical challenges at my job with a global consumer products corporation during the week. I enjoy analyzing consumer shopping behaviors to further strengthen the mental and physical ability of our brands. My dream is to own my own consulting business one day. If you're a go-getter looking for an adventure and up for some witty banter, send me a message!
Both Adeline and Declan have busy schedules and need to find a time to meet. Here is a glimpse of at their calendars for the next 7 days.

Adeline's Schedule

Declan's Schedule
A4: City Descriptions

**The City of Alpha**

<table>
<thead>
<tr>
<th>City</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>Phi</td>
</tr>
<tr>
<td>Area (sq mi)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4,484.14</td>
</tr>
<tr>
<td>Land</td>
<td>4,359.81</td>
</tr>
<tr>
<td>Water</td>
<td>124.33</td>
</tr>
<tr>
<td>Elevation</td>
<td>2,651 ft</td>
</tr>
<tr>
<td>Population (2010)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12,968</td>
</tr>
<tr>
<td>Density</td>
<td>3/sq mi</td>
</tr>
</tbody>
</table>

**Description**

Alpha is a quiet city in Phi County that prides itself as a close-knit community with a good school system. A post office, grocery store, sporting goods store and gas station are the only businesses within city limits. It is also the home of the Phi National Park, known for its great hiking and biking trails. The city of Alpha is a great place to live if you want to get away from the hustle of bustle of city life and enjoy the great outdoors, but expect to pay a premium for homes in this area due limited availability and high demand.

**The City of Delta**

<table>
<thead>
<tr>
<th>City</th>
<th>Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>Delphia</td>
</tr>
<tr>
<td>Area (sq mi)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8,782.59</td>
</tr>
<tr>
<td>Land</td>
<td>8,673.29</td>
</tr>
<tr>
<td>Water</td>
<td>109.30</td>
</tr>
<tr>
<td>Elevation</td>
<td>128 ft</td>
</tr>
<tr>
<td>Population (2010)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5,283</td>
</tr>
<tr>
<td>Density</td>
<td>1.1/sq mi</td>
</tr>
</tbody>
</table>

**Description**

Delta is a large farming community in the county of Delphia. There is very little commercialization in this area but is the primary source of beef, chicken, and agricultural for the country. Delta is also the location of the only StarRail station, the railway system used for travel between different countries, in the country of Zen. Many Zen residents travel through Delta to access the StarRail but never leave the train station.
The City of Beta

City: Beta  
County: Apollo

Area (sq mi)  
- Total: 1,603.31  
- Land: 1,588.11  
- Water: 14.20

Elevation: 943 ft

Population (2010)  
- Total: 6,497  
- Density: 4.1/sq mi

Description
Beta was established many years ago as a lumber town, with the Beta Lumber Company as the primary employer there. With little interest from residential developers over the years, this city became a booming industrial town. Beta is home to many companies such as Demeter Food Production, Athena Automobiles and Apollo Gas and Electric. While some residential property exits in this city, most of the workers at the plants come in by train to work each day.

Dining Options
Lumberjacks
Address: 31 Sonet St  
Phone: 555-0686  
Hours: M-F, 6:00 am - 2:00 pm  
Description: Cozy eatery serving up traditional breakfast and lunch Monday - Friday.

Reviews
4.0 ★★★★★  9 Google reviews

Ted's Café
Address: 34 Sonet St  
Phone: 555-1212  
Hours: M-S, 11:00 am - 11:00 pm  
Description: Burgers, fried fish sandwiches and other pub-grub classics go alongside a beer at this longtime hangout.

Reviews
4.2 ★★★★★  40 Google reviews
The City of Epsilon

City: Epsilon
County: Zeus

Area (sq mi)
  • Total: 6,337.05
  • Land: 2,647.56
  • Water: 3,689.49

Elevation: 306 ft

Population (2010)
  • Total: 43,575
  • Density: 16.5/sq mi

Description
Epsilon sits on Lake Zeus and attracts heavy tourism during the summer months. While the Lake Zeus Resort is the largest offers the most luxurious accommodations on the water, private vacation rentals and bed and breakfasts are also available. Visitors will enjoy an abundance of outdoor activities such as jet ski and speed boat rentals, water slides, mini-golf and more. In addition to the fun activities, Epsilon boasts award-winning restaurants and is the home of the Epsilon Brewery.

Dining Options
Epsilon Brewery
Address: 2501 Monarch St
Phone: 555-2739
Hours: Tues-Sun, 11:00 am - 11:00 pm
Description: Taproom located in historic bottling plant offering house-brewed beers, ping-pong, and TVs.
Reviews: 4.2 ★★★★★ 47 Google reviews

Bobbyquivaris
Address: 1450 Lombard St
Phone: 555-8880
Hours: Sun-Sat, 4:00 pm - 11:00 pm
Description: Clubby, upscale fixture for steaks, chops, & seafood in a lively, jackets-suggested setting.
Reviews: 4.3 ★★★★★ 43 Google reviews
The City of Sigma

City: Sigma
County: Sentinel

Area (sq mi):
- Total: 1,301.82
- Land: 1,287.82
- Water: 14.20

Elevation: 103 ft

Population (2010):
- Total: 6,497
- Density: 5/sq mi

Description:
Sigma is the central business district in Zen and home of the Zen International Airport. The downtown district is filled with the country's finest law firms, insurance agencies, financial advisors and other professional organizations. Housing is available just outside the city, but many professionals keep apartments in the downtown district during the week and head home to other cities on the weekend.

Dining Options

The Sentinel
Address: 37 New Montgomery St
Phone: 555-9950
Hours: M-F, 6:00 am - 2:00 pm
Description: Low-key café offering simple options breakfast, brunch, and lunch. Perfect for a business meeting or sharing a meal with a friend.

Galette 88
Address: 88 Hardie Pl
Phone: 555-9119
Hours: M-F, 11:00 am -11:00 pm
Description: Creative cocktails and cuisine are offered at this upscale restaurant close to the airport.

Mozzeria
Address: 3228 Divisardo St
Phone: 555-9410
Hours: M-F, 11:00 am -10:00 pm
Description: Neapolitan pies are wood-fired and served with antipasta and craft cocktails in the basement of the basement of the Titan Tower.
The City of Gamma

City: Gamma
County: Nu

Area (sq mi)
- Total: 3,310.07
- Land: 1,867.39
- Water: 1,442.68

Elevation: 226 ft

Population (2010)
- Total: 34,229
- Density: 18.3/sq mi

Description
Gamma is a small lake front community just outside the big city of Upsilon. Lake Artemis provides a beautiful backdrop for the lakefront chalets and the breathtaking Sugar Hill Golf Course. Unlike the city of Upsilon on Lake Zeus, Gamma is primarily residential with minimal tourism. Gamma is a popular day trip destination with it's public beaches, local wineries, premier golf courses, and award-winning cuisine.

Dining Options

El Habanero
Address: 55 Webster St
Phone: 555-2339
Hours: Tues-Sun, 11:00 am - 11:00 pm
Description: Mexican cantina with outdoor seating, festive cocktails, and award winning cuisine
Reviews
4.7 ★★★★★ 574 Google reviews

Fog Harbor Fish House
Address: 1914 Fillmore St.
Phone: 555-1101
Hours: Sun- Sat, 11:00 am -11:00 pm
Description: Seafood dominates the menu at this lakeside, multi-level restaurant with a lively happy hour.
Reviews
4.0 ★★★★★ 58 Google reviews

Grimaldi's
Address: 3636 McKinney Ave
Phone: 555-4611
Hours: Mon- Sat, 11:00 am - 10:00 pm
Description: Zen-style stuffed pizza and thin-crust pies plus other Italian dishes in a casual setting.
Reviews
4.2 ★★★★★ 200 Google reviews
The City of Upsilon

City: Upsilon
County: Lethe

Area (sq mi)
- Total: 478.70
- Land: 478.29
- Water: 0.41

Elevation: 397 ft

Population (2010)
- Total: 58,776
- Density: 208.5/sq mi

Description:
Upsilon is the most populous city in Zen and is located in the south central part of the country. It has significant influence on the country's financial position as it comprises the offices and headquarters for all major technology, fashion, retail, and healthcare corporations. While the financial district primarily resides in Sigma, many financial institutions also have offices in Upsilon. It is also the cultural, entertainment, and education capital of Zen. Upsilon is the home of the Silverbacks baseball team, the Apollo Theater, the famed Demeter Culinary School, and Athena University.

Dining Options

**Bin 36**
Address: 339 Dearborn St.
Phone: 555-5463
Hours: Mon-Sat, 4:00 pm - 3:00 am
Description: Casual spot serving New Zen fare, cheese plates, wine flights in a contemporary open space.
Reviews: 4.1 [5 Google reviews]

**Zappia's**
Address: 108 Kinzie St
Phone: 555-9555
Hours: Mon-Sat, 4:00 pm - 3:00 am
Description: Rustling, fl level space where groups go for small and large plates plus carafes of wine.
Reviews: 4.3 [124 Google reviews]

**Marni Thai**
Address: 1249 Madison Rd.
Phone: 555-8963
Hours: Sun-Sat, 4:00 pm - 11:00 pm
Description: Long running restaurant serving a traditional Thai menu in a compact and unassuming setting.
Reviews: 3.7 [283 Google reviews]
The City of Iota

City: Iota
County: Ioke

Area (sq mi)
- Total: 834.90
- Land: 815.60
- Water: 19.30

Elevation: 176 ft

Population (2010)
- Total: 3,293
- Density: 4/sq mi

Description
Iota is the smallest city in Zen and location of the most expensive real estate in the country. Sprawling estates dot the countryside and provide the wealthy land owners their escape from the big city life of Upsilon.

The City of Kappa

City: Kappa
County: Kakia

Area (sq mi)
- Total: 1,792.89
- Land: 1,774.61
- Water: 18.28

Elevation: 43 ft

Population (2010)
- Total: 5,779
- Density: 3.3/sq mi

Description
Kappa is a small mining town on the south west side of Zen. It is also home to Zen’s largest power plant. Residential areas exist in the areas surrounding the mines and plant but primarily remains an undeveloped part of the country.
The City of Omega

City: Omega
County: Orion

Area (sq mi)
- Total: 2,258.19
- Land: 2,243.99
- Water: 14.20

Elevation: 429 ft

Population (2010)
- Total: 9,263
- Density: 4.1/sq mi

Description
Omega is the southern most city in Zen. Omega is best known for their historic homes and sprawling horse farms. For 51 weeks of the year, Omega is a quiet town devoid of commerce and without much traffic. But for the 1st week of May each year, many of the citizens of Zen and surrounding countries travel to Omega for the Zen Derby. This annual horse racing event is over 100 years old and a cherished tradition held at Omega Downs.

Dining Options
Thoroughbred Fare
Address: 1722 Rose Blvd
Phone: 555-6001
Hours: M-F, 6:00 am - 2:00 pm
Description: Quaint café with a menu of homestyle Southern breakfast and lunch eats in the tiny bustling restaurant.
Reviews: 3.7 ★★★☆☆ 45 Google reviews

Triple Crown
Address: 433 Market St.
Phone: 555-6569
Hours: M-S, 11:00 am - 11:00 pm
Description: Substantial breakfasts & burgers, plus bourbon and happy hours are offered at this quirky, historic location in horse country.
Reviews: 4.1 ★★★★★ 37 Google reviews
A5: Train Schedule

Zen Train Schedule and Costs

<table>
<thead>
<tr>
<th>Line</th>
<th>Round Trip</th>
<th>One Way</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>$14.00</td>
<td>$7.50</td>
</tr>
<tr>
<td>Purple</td>
<td>$9.00</td>
<td>$5.50</td>
</tr>
<tr>
<td>Orange</td>
<td>$10.00</td>
<td>$5.50</td>
</tr>
</tbody>
</table>

Green Line
*Trains depart every 20 minutes*

<table>
<thead>
<tr>
<th>City</th>
<th>First Train</th>
<th>Last Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>4:45a</td>
<td>1:30a</td>
</tr>
<tr>
<td>Beta</td>
<td>4:55a</td>
<td>1:40a</td>
</tr>
<tr>
<td>Epsilon</td>
<td>4:23a</td>
<td>1:59a</td>
</tr>
<tr>
<td>Sigma</td>
<td>5:00a</td>
<td>1:50a</td>
</tr>
<tr>
<td>Delta</td>
<td>5:00a</td>
<td>1:45a</td>
</tr>
<tr>
<td>Gamma</td>
<td>5:10a</td>
<td>1:55a</td>
</tr>
<tr>
<td>Upsilon</td>
<td>4:05a</td>
<td>2:00a</td>
</tr>
</tbody>
</table>

Purple Line
*Trains depart every 15 minutes*

<table>
<thead>
<tr>
<th>City</th>
<th>First Train</th>
<th>Last Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>4:30a</td>
<td>1:30a</td>
</tr>
<tr>
<td>Kappa</td>
<td>4:45a</td>
<td>1:45a</td>
</tr>
<tr>
<td>Omega</td>
<td>4:30a</td>
<td>1:30a</td>
</tr>
<tr>
<td>Iota</td>
<td>4:45a</td>
<td>1:45a</td>
</tr>
<tr>
<td>Upsilon</td>
<td>4:30a</td>
<td>1:30a</td>
</tr>
<tr>
<td>Gamma</td>
<td>4:45a</td>
<td>1:45a</td>
</tr>
</tbody>
</table>

Orange Line
*Trains run 24 hours a day, 7 days a week*

<table>
<thead>
<tr>
<th>City</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma &amp; Gamma</td>
<td>every 30 min, on the hour and every half of the hour</td>
</tr>
<tr>
<td>Delta &amp; Upsilon</td>
<td>every 30 min, at quarter after the hour, and quarter till the hour</td>
</tr>
</tbody>
</table>
A6: Student Survey Instructions

SURVEY INSTRUCTIONS

1. PLEASE SELECT A SEAT IN THE ROOM

2. OPEN UP YOUR LAPTOP, CONNECT YOUR HEADPHONES, & OPEN A WEB BROWSER
   (e.g. Google Chrome, Internet Explorer)

3. IN THE ADDRESS BAR OF YOUR WEB BROWSER, TYPE THE FOLLOWING LINK
   
   **SURVEY LINK** → http://is.gd/UuBPL4

4. YOUR SURVEY PASSWORD
   
   **SURVEY PASSWORD** → VOL452JS
A7: Mechanical Turk “Attention Check”

Population (2020)
- Total: 5,779
- Density: 3.2/sq mi

Description
Kappa is a small mining town on the south west side of Zen. It is also home to Zen's largest power plant. Residential areas exist in the areas surrounding the mines and plant but primarily remains an undeveloped part of the country.

If you are paying attention, answer "Strongly Agree" to this question.
- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

The City of Gamma

<table>
<thead>
<tr>
<th>City</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>Nu</td>
</tr>
</tbody>
</table>

Area (sq mi)
- Total: 3,310.07
- Land: 1,967.94
A8: Mechanical Turk End of Survey Quality Questions

Please answer the following questions. Your responses to these questions DO NOT affect your pay for this task. We simply want to collect some information to help us gain a better understanding of your environment while taking our survey. Thank you!

On a scale from 0-100, how much effort did you put into taking this survey?
- 0 = I just clicked through the screens and didn’t really read the information
- 50 = I read through the information on the screens but didn’t really pay attention to the details
- 100 = I carefully read through the information presented, took notes when necessary, and came up with the best answer

On a scale from 0-100, what was the noise level around you while you took this survey?
- 0 = very quiet, you could hear a pin drop
- 25 = fairly quiet, low talking or light music
- 50 = about average
- 75 = fairly loud, like a busy restaurant at lunch
- 100 = very loud, like a rock concert or busy construction site

Describe the type of location where you took this survey. (Office, public park, library, subway, etc...)

Were you interrupted at all while taking this survey?
(After a phone call, quick response to a text/email, check Facebook, watch a commercial, etc...)

- Yes
- No

NEXT
APPENDIX B: Emergency Department Experiment Screen Shots

B1: Problem introduction and description

Thanks for agreeing to participate in this study evaluating Epic usability! Below is a description of what you will be doing during this study.

1. First, on the next page you will listen to an audio clip describing a patient arriving via air care.
2. Second, you will have the opportunity to review this patient’s historical record in the Epic shadow environment. (View only format used during Epic downtime.)
3. Last, you will be provided a place to briefly document the patient's anticipated clinical course and what diagnostic and therapeutic interventions you anticipate will be necessary in the first hour of care.

B2: Aircare audio clip

Please click on the audio clip below to listen to a description of the patient arriving via aircare.

Audio Clip

NEXT
B3: Epic Instructions

Please click "Alt-Tab" to begin logging in to the Epic Maintenance Read Only environment to review the medical record of the arriving patient.

STEP 1:

STEP 2: Start Log in as usual

STEP 3: BUT, instead of going into Epic Hyperspace - click on the Epic Downtime folder.

STEP 4: Next, click on Hyperspace Maintenance ReadOnly

STEP 5: Please review the medical record of the arriving patient.

Medical record: XYZ
B4: This is a sample screen shot provided by the training department to avoid any HIPAA violations

B5: This is the screen where the physicians enter their medical decision making.
B6: Information and scoring sheet for quality scoring of participant decision making.

MEDICAL DECISION MAKING

DESCRIPTION OF THIS EXPERIMENT:
This was a computer-based, laboratory experiment conducted with emergency department physicians (n=24). First, the ED physicians reviewed a description of the experiment and provided consent to participate. Second, they listened to an audio clip, recorded by one of their ED physicians, describing a hypothetical patient arriving via aircare. Third, they were provided the medical record number for the arriving patient and instructed to access EPIC and review the patient’s historical medical record. Last, they typed responses to a series of questions geared toward understanding their proposed plan of care for the arriving patient.

AUDIO CLIP PRESENTED TO PHYSICIAN PARTICIPANTS:
“AirCare 1 will be en route with a 43-year-old male restrained driver in a motor vehicle accident who is agitated on scene and was unable to be controlled and intubated for airway protection. He is notable to have a heart rate of 130 and blood pressure 150 over 87. A respiratory heart rate of 28, setting 99%, his breath sounds are clear bilaterally, his belly is soft, he has no obvious extremity trauma. He is noted to be somewhat rigid after the intubation. Of note, his wife did mention that he has a history of allergies to anesthesia. ETA is approximately 5 minutes. Any questions?”

IMPORTANT INFORMATION ABOUT THIS PATIENT/THIS ACCIDENT:
• This is a low-mechanism MVA with a low probability of severe injury.
• The patient was agitated on scene which is most likely due to alcohol intoxication and/or TBI/ICH. He has a history of alcohol related injuries per a prior ED visit.
• Succinylcholine is the depolarizing paralytic agent used by aircare. The patient has a history of Malignant Hyperthermia as noted in his medical record.
• No history of diabetes, cardiovascualr disease, seizures, or neurological issues.

QUESTIONS POSED TO EACH PHYSICIAN PARTICIPANT:
Assume no changes in the patient’s clinical exam. Please answer the following questions below:
1. What are the top items in your differential?
2. What is anticipated clinical course for this patient in the first hour?
3. What are the key factors in your responses?

TASK: SCORING THE RESPONSES
You will be asked to score each of the 24 responses on a scale from 0-100 pts. Each question provides some guiding information.
SCORING SHEET

Score A. Quality of the physician response to possible causes for the patient’s altered mental status/agitation on:

Things to consider:
- This is a low-mechanism MVA with a low probability of severe injury
- A response that receives a perfect score must include both the possibility of TBI/ICH related to MVA or intoxication in their differential

Score B. Quality of the physician response in terms of addressing the rigidity of the patient post intubation.

Things to consider:
- A response that receives a perfect score must note the possibility of Malignant Hyperthermia

Score C. Overall quality of the physician decision making.

Things to consider:
- Bayesian Decision Theory: What is the cost if they make the wrong decision?
  - When considering the top items in their differential, a response that achieves a perfect score should appropriately balance the risks and benefits of the items with the highest probability.
  - When considering the anticipated clinical course for the first hour, a response that achieves a perfect score must include a trauma stat and appropriately address all of the items in their differential.

Score D. Overall clarity and thoroughness of the physician response.

Please use the following scale to score each physician response, assigning from 0 to 100 pts.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pts</td>
<td>Inacceptable</td>
</tr>
<tr>
<td>25 pts</td>
<td>Acceptable</td>
</tr>
<tr>
<td>50 pts</td>
<td>Above Average</td>
</tr>
<tr>
<td>75 pts</td>
<td>Excellent</td>
</tr>
<tr>
<td>100 pts</td>
<td>Perfect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physician Participant</th>
<th>Score A</th>
<th>Score B</th>
<th>Score C</th>
<th>Score D</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_9RJKvxvu0pmOgd</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_9AAPs3ySOhhONJx</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_bq1LK3plK0tOTh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_7JUQG3829Wz0tBF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_08p1E4i87K126N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_DnypOL2G9OJIV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_3AW8pWCz9F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_22WapwCeYrUref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_gYUOT37L51h6SN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


interdependence and supervisory experience on management assessments of
resource planning systems. *Production and Operations Management* 17(1) 93-
106.

10.Bendoly, E., 2011. Linking task conditions to physiology and judgment errors in

reawakening practical issues in research. *Behavioral Issues in Operations
Management* (pp. 1-22). London: Springer.

12.Black A.D., Car J., Pagliari C., Anandan C., Cresswell K., Bokun T., Procter, R.,
Majeed, A., Sheikh, A., 2011. The impact of eHealth on the quality and safety of

laboratory investigation of the role of experience and feedback. *Manufacturing &
Service Operations Management* 10(3) 519-538.


Adaptive learning in a laboratory experiment. *Manufacturing & Service
Operations Management* 10(4) 590-608.

16.Botvinick, M., Braver T., 2015. Motivation and cognitive control: from behavior to
neural mechanism. *Psychology* 66(1) 83.


