Spring 6-10-2012

The Effects of Diagrams and Relational Complexity on User Performance in Conditional Probability Problems in a Non-Learning Context

Vincent J. Kellen
DePaul, vkellen@uky.edu

THE EFFECTS OF DIAGRAMS AND RELATIONAL COMPLEXITY ON USER PERFORMANCE IN CONDITIONAL PROBABILITY PROBLEMS IN A NON-LEARNING CONTEXT

By

Vince Kellen

A dissertation submitted to the School of Computing,

College of Computing and Digital Media of DePaul University

In partial fulfillment of the requirements for the

Degree of

Doctor of Philosophy

June 10, 2012

ABSTRACT

Many disciplines in everyday life depend on improved performance in conditional probability problems. Most adults struggle with conditional probability problems and several prior studies have shown participant accuracy is less than 50%. This study examined user performance when aided with computer-generated Venn and Euler type diagrams in a non-learning context. Despite the prevalence of research into diagrams and extensive research into conditional probability problem solving, this study is one of the only studies to apply theories of working memory to predict user performance in conditional probability problems with diagrams. Following relational complexity theory, this study manipulated problem complexity in computer generated diagrams and text-only displays to improve user performance and perceptions of satisfaction. Partially consistent with the study hypotheses, complex visuals outperformed complex text-only displays and simple text-only displays outperformed complex text-only displays. However, a significant interaction between users' spatial ability and the use of diagram displays led to a degradation of low-spatial user performance in the diagram displays when compared to high spatial users. Participants with less spatial ability were significantly impaired in their ability to solve conditional probability problems when aided by a diagram.

APPROVAL

DePaul University
College of Computing and Digital Media

Dissertation Defense Report

I have read the dissertation written by:

Name Vince Kellen SSN ID: 0195746

(***To the advisor)***: The following dissertation title is identical to the one on the title page of the draft returned to the student. This title is approved by me and it is to be used when the final copies of the dissertation are prepared.

Title of dissertation:
The Effects of Diagrams and Relational Complexity on User Performance in Conditional Probability Problems in a Non-Learning Context

Advisor's Initials S. C.

☒ Acceptable. Candidate may proceed to make final copies

☐ Pass, with revisions stated below:

☐ Not Acceptable. Please explain:

Print Name	Signature	Date
Susy S. Chan Advisor (Print Name)	Susy S. Chan	5/25/2012
Xiaowen Fang 1st Reader (Print Name)	[signature]	5/25/2012
Daniela Raicu 2nd Reader (Print Name)	Daniela Raicu	5/29/2012
Ken Calvert 3rd Reader (Print Name)	Kenneth L. Calvert	5/29/2012
4th Reader (Print Name)	Signature	Date

ACKNOWLEDGEMENTS

A dissertation is hardly the work of just one person. I would like to thank Dr. Xiaowen Fang for his keen insights, help and direction and more importantly, Dr. Susy Chan for having the good sense to engage me two decades ago when I attended one of her excellent graduate courses. She has been mentor for and a colleague with me in so many ways throughout the years. Dr. Fang and Dr. Chan are exemplars of the best of what the academy offers students and society. I am honored to have had the chance to benefit from their intellect, wisdom, guidance and patience. I would like to thank Dr. Scott Scarborough for the ongoing prodding and encouragement throughout the process. I would also like to give some extra thanks to my two children, J.R. and Claire for whom I started this endeavor. Without them I could not have finished. Finally, I would like to thank my wife Liz. Success here, or for that matter anywhere, was not possible without her perpetual support, patience and faith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. RESEARCH PROBLEMS, OBJECTIVES AND SIGNIFICANCE

With the explosion of availability of information in the past few decades, people are inundated with numbers. News coverage about health, security and safety risks often require people understand quantitative problems and make sound inferences. Understanding the chances of contracting diseases, weighing the pros and cons on the risks of driving or flying or understanding election results requires good probability and quantitative reasoning skills. Researchers are finding many people lack these skills and make erroneous judgments.

With a greater reliance on web and mobile interactions for complex quantitative data, human computer interface (HCI) designers will need to pay particular attention to how information displays may or may not help. While the use of visuals is generally thought to help in problem solving, some common quantitative problems seem to defy reliable improvement with visuals. Problems which require people to process multiple relationships between terms and incorporate math within their reasoning are especially prone to error. Probability and statistical reasoning problems are a prime example of these kinds of problems. Prior research that tries to explain why people err and how to make improvements in visual displays are conflicting (Brase, 2008; Evans, Handley, Perham, Over & Thompson, 2000; Girotto & Gonzalez, 2001; Johnson-Laird, Legrenzi, Girotto. & Legrenzi, 1999; Mellers & McGraw, 1999; Stenning & Oberlander, 1995; Sloman, Over, Slovak & Stibel, 2003). In some cases, a visual representation seems to help performance, in other cases it hurts.

Many disciplines in everyday life depend on improved performance in statistical and probability reasoning problems. In tasks such as disease diagnosis, interpretation of evidence in criminal trials and management of security and risk data, people need to process statistical and probability data, relying on syllogistic, conditional reasoning and math skills, to make critical inferences and decisions. Computer-aided displays are now nearly ubiquitous in today's work and home environments. For HCI designers, understanding what factors improve performance is important.

According to theories of working memory (Baddeley and Hitch, 1974; Baddeley, 1986; Baddeley, 2007; Miyake & Shah, 1999) and mental models research (Johnson-Laird et al., 1999; Fangmeier, Knauff, Ruff & Sloutsky, 2006; Knauf, Jola & Strube, 2001; Ruff, Knauff, Fangmeier, Spreer,, 2003; Vandierendonck & De Vooght, 1997), visual displays should aid in these tasks as they can make better use of working memory, reduce cognitive load and facilitate mental model construction, thus generating better judgments. The conditional probability problem is a common and difficult quantitative reasoning task. While it can be solved using a complex formula (developed by Thomas Bayes in 1764), most people have not learned Bayes' theorem. Solving the problem requires people to build a mental model of relationships between elements in the problem. Individuals' working memory can be overloaded. While a significant amount of research has examined why people make errors in conditional probability reasoning problems (Eddy, 1982; Gigerenzer & Hoffrage, 1995; Sloman, et al., 2003), very little of it has looked systematically at optimizing people's performance using visuals. Moreover, practically no research has applied theories of working memory to guide facilitation of these kinds of problems. Individuals do vary in their working memory capacities and understanding the limits of working memory capacity may matter when constructing visuals.

Solving these kinds of problems requires all aspects of working memory, including attentional control, linguistic and visual (object and spatial) processing. Visual displays can engage both automatic perceptual and more effortful cognitive mechanisms as people process the external representation (display) and construct a mental model that depicts the problem state. It is quite possible that individuals cannot effectively process a visual and solve the problem at the same time, especially if processing the visual representation competes for limited working memory resources. Some diagrams commonly thought to help statistical reasoning problems that are widely used include Venn and Euler diagrams.

Counter-intuitively, some research is finding that Venn and Euler type diagrams, which try to depict set inclusions and conditional relationships directly, may actually hurt performance (Brase, 2008; Calvillo, DeLeeuw & Revlin, 2006; Sloman et al., 2003). Mental models theory (Johnson-Laird et al., 1999), relational complexity theory (Halford, Wilson & Phillips, 1998), and explanations of working memory limits (Cowan, 2000; Cowen et al., 2005; Rouder et al., 2008) can provide guidance for designing visuals to help people solve these problems. Despite the preponderance of research in each of these domains and into the problems with performance in statistical reasoning, surprisingly little has been brought to bear on improving performance in statistical reasoning problems with visuals in non-learning conditions.

By applying theories of working memory, mental model theory and relational complexity theory this study examines how designers can construct visuals to help in these kinds of problems. This study's objective is to more precisely identify how Venn and Euler type diagrams can be constructed to help performance on these problems. This study will examine how manipulations of the problem complexity (as described by the number of relations or entities that must simultaneously be kept in the focus of attention) and using visuals in place of text where possible can improve performance. As research into relational complexity and limitations of working memory indicate, when problems are presented in such a way as to keep the number of simultaneously interacting elements in working memory well within the limit of about four items, people's performance should improve. Prior research may have overlooked the significance of this critical working memory limit in solving these kinds of problems. Additionally, when displays selectively use visuals in place of text, demands on the attentional capacity that has to coordinate text processing, visual processing and mental model construction are reduced and modality-specific forms of short term memory capacity might be better leveraged, which in turn should contribute to improved performance for the individual.

Because the potential relationships between elements in the visual display are not clear and can easily exceed three or four items that must be held in memory simultaneously, these problems most likely lie at and beyond the edge of working memory capacity. Perhaps some of the conflict and poor performance in prior research may be because displays used in those studies contain too many elements or relationships between elements and do not carefully help the user manage limited working memory resources by leveraging perceptual processes to guide attention and sequence the construction of the mental model. Also, by recruiting younger participants from universities, many prior studies may have overlooked the marked decrease in spatial processing abilities (which are a critical working memory component implied in reasoning skills) that naturally occurs between the people between the ages of 20 and 45 (Hale, Myerson, Emery, Lawrence & Dufault, 2008). Everyday situations of problem comprehension involve people of varying working memory capacities.

For visual interface designers, this study will identify clearer design guidelines that, if followed, may help users make better judgments in critical conditional probability and other similar quantitative reasoning problems. This study intends to provide a theoretical explanation and validate that using visuals that carefully control relational complexity and conserve working memory will help people understand the problem.

CHAPTER 2. LITERATURE REVIEW

The literature into the role of visuals in reasoning problems is diverse and involves many disciplines. While some limited research within the human-computer interaction (HCI) and computer science disciplines does exist, a large body of prior research exists in cognitive psychology, neuropsychology, economics, and learning and instruction research.

This study starts by examining statistical reasoning problems (specifically conditional probability problems), discusses the causes and suggested cures for poor performance, shows how statistical reasoning relates to other forms of reasoning and how neurological evidence may be converging on explanations for performance. Finally, this review of the literature examines three theories that interrelate and are relevant for this study: theories of working memory, relational complexity theory and cognitive load theory.

PERFORMANCE OF CONDITIONAL PROBABILITY PROBLEMS

With the vast amount of complex information around us, statistical reasoning skills are critical for both the layman and the professional. The medical, social science, economics, legal and risk management disciplines use statistical problem solving in contexts that require many people who are unskilled in statistics and probabilities to make critical judgments and decisions. Everyday contexts include healthcare professional communicating disease diagnosis and treatment, lawyers, jurors and court officials' interpretation of evidence in criminal and civil cases, business professionals' decision-making regarding risks and probabilities of future events (Gigerenzer & Hoffrage, 1995; Kurzenhauser & Hertwig, 2006; Boumans, 2008; Kahnamen & Lovallo, 1993; Sedlmeier, 2000). More and more, the news media, blogs and social media disseminate poll data, social science surveys and many scientific studies requiring people to have statistical literacy.

Many people seem to lack good statistical reasoning skills creating a current context that some have termed “innumeracy” which refers to peoples’ inability to reason with numbers and mathematical concepts (Paulos, 1988). Several studies show participants’ accuracy on conditional probability problems is low to intermediate, ranging from 6% of the problems correct to 62% depending on the textual format (Brase, Cosmides, Tooby, 1998; Gigerenzer & Hoffrage, 1995).

Different verbal formats for conditional probability problems

An example of a conditional probability problem, shown below, has been used in several studies (Gigerenzer & Hoffrage, 1995):

> The probability of breast cancer is 1% for women at age forty who participate in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

The normative answer requires applying a form of Bayes’ theorem to calculate the posterior probability that a woman with positive results has cancer. The symbols *H* and *–H* denote the two hypotheses (cancer and no cancer, respectively) and *D* denotes the data representing the positive test results. The formula for computing $p(H \mid D)$ is as follows.

$$p(H \mid D) = \frac{p(H)\,p(D \mid H)}{p(H)\,p(D \mid H) + p(-H)\,p(D \mid -H)} = \frac{(.01)(.80)}{(.01)(.80) + (.99)(.096)} = .078$$

The method for calculating conditional probabilities was developed by Thomas Bayes in 1764. Only those who have studied Bayes’ theorem well would have knowledge of how to apply it to these kinds of problems. Prior studies show that even physicians estimate probabilities incorrectly when presented in text form. Eddy (1982) shows that 95% of the physicians queried estimated the probability of $p(H \mid D)$ in this problem between 70-80% not 7.8%. Other prior research indicates subject accuracy is typically poor for this text-only, probability representation of the problem (Gigerenzer & Edwards, 2003; Gigerenzer & Hoffrage, 1995; Girotto & Gonzalez, 2001; Kurzenhauser & Hoffrage, 2002; Sedlmeier, 2000; Sloman, et al., 2003).

Gigerenzer and Hoffrage (1995) have argued that expressing conditional probability problems as *natural frequencies* rather than as probabilities improves subjects' performance. Natural frequencies arise from natural sampling and are not normalized with respect to base rates (Gigerenzer & Hoffrage, 1999). Rather, natural sampling mimics the process of encountering instances in a population sequentially – in other words, as human beings normally encounter observations in their lives. Cosmides and Tooby (1996) and Gigerenzer & Hoffrage (1995) propose that human evolution has endowed us with minds that can reason with natural frequencies better than probabilities. Therefore participants should perform better when probability problems are expressed as natural frequencies.

A natural frequency version of a conditional probability problem, following Gigerenzer & Hoffrage (1995), is as follows:

> 10 out of every 1,000 woman at age forty who participate in routine screening have breast cancer. 8 out of every 10 women with breast cancer will also get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography. Here is a new representative sample of 100 women at age forty who got a positive mammography in a routine screening. How many of these women do you expect to actually have breast cancer?

Natural frequencies can also be depicted in a contingency table, shown below.

	H	Not H	
D	8	95	103
Not D	2	895	897
	10	990	1000

While natural frequencies make clearer the problem's hidden set structure, so do other problem formats, which can depict the quantitative information as probabilities rather than frequencies (Evans, et al., 2000; Girotto & Gonzalez, 2001; Johnson-Laird et al., 1999; Mellers & McGraw, 1999). Displays that clarify the problem's set structure help people solve these problems. This particular problem's nested set structure includes a) women who are regularly screened, b) women who test positive for the disease and c) women who have the disease.

In addition, a natural frequency representation has a simpler mathematical form of Bayes' theorem:

$$p(H \mid D) = \frac{d \,\&\, h}{d \,\&\, h + d \,\&\, {-h}} = \frac{8}{8+95} = 0.078$$

This format has fewer elements and fewer computations needed to calculate an answer.

Conditional probability reasoning and other forms of reasoning

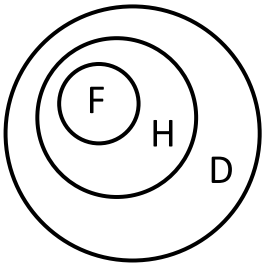
Conditional probability problems share similarities with syllogistic, conditional and relational reasoning problems, all of which have been extensively researched. To solve conditional probability problems, people have to determine what are the relevant categories or sets and determine how these sets relate to each other. For example, the following is a syllogistic problem:

Is the following argument valid?

All F are H
All H are D

Some F are not D

These problems can be easily depicted as an Euler diagram:



In this case, the set-subset structure between the elements F, H and D, are depicted clearly in the diagram. Conditional probability problems are frequently represented in Euler diagrams for the same purpose: to visually depict the relationships between the sets represented in the text of the problem.

Conditional reasoning (which is related, but distinct from solving conditional probability problems) involves drawing inferences from problems in which the occurrence of an event depends on the occurrence of another event. An example is “If the flashlight runs out of battery, then it won’t shine.” In a similar fashion, conditional probability (statistical reasoning) problems require people to infer a probability of one event that depends on the probability of a prior event, e.g., if the person tests positive for a disease, they may actually have the disease.

Relational reasoning problems require people to draw inferences based on relations between items:

A is bigger than B

B is smaller than C

A is smaller than C

D is smaller than B

Is A smaller than D?

Relational reasoning is needed to determine the sequence of items in an array and is brought to bear to resolve relationships based on size, width, weight, importance, position or any relational property. Conditional probability problems involve relational reasoning if the problem solver links numeracy to the set inclusion, e.g., the population of those with the disease is smaller than the population of those who tested positive for the disease, which in turn, is smaller than the population of all those tested for the disease. Conditional probability problems involve more than just syllogistic reasoning. People need to comprehend and integrate set size and mathematical operations based on the set relationships. In this regard, “thinking about numbers” may be interacting with “thinking about sets” to create additional complexity beyond syllogistic reasoning.

USE OF VISUALS IN CONDITIONAL PROBABILITY PROBLEM SOLVING

While ample research over the past few decades has examined the errors people make in performing statistical reasoning, a lesser but growing body of research has addressed how to teach people to solve statistical problems (Gigerenzer & Hoffrage, 1999; Kurzenhauser & Hoffrage, 2002; Sedlemeier, 2000). Unfortunately, many everyday settings that require statistical reasoning abilities involve people who haven’t been taught how to reason with statistics or have failed to retain the associated skills. Everyday settings are also not normally conducive for teaching numeracy and statistical reasoning skills and may require visuals to help (Nelson, Reyna, Fagerlin, Lipkus, Peters, 2008).

Intuitively, one would think that using diagrams to depict clearly the set structure of the problem would help performance. Diagrams are often employed to facilitate thinking processes in problems that are too complex to be solved exclusively with internal cognition (Larkin & Simon, 1987; Hoffman, 2007). Diagrams can be considered an analogical representation with a structure that models the semantics of the problem (Kulpa, 1994).

One common diagram, an Euler diagram, failed to provide significant improvement in either a probability or frequency format (Sloman et al., 2003) and performed worse than iconic displays (Brase, 2008) which depict each item in the frequency of observations as an icon. According to Brase (2008), iconic displays are superior to Venn diagrams because they better approximate ‘actual ecological presentations.’ In other words, in the ecological rationality approach he advocates, iconic displays are similar to natural frequency formats. These iconic representations more closely resemble naturalistic settings in which people encounter events or cases as a sequence of occurrences over time.

Doctors’ comprehension of statistical information can be improved with this natural sampling approach (Hoffrage, Gigerenzer, Krauss, Martignon, 2002). In a study examining deductive reasoning, Calvillo, DeLeeuw and Revlin (2006) found a similar disadvantage for Euler diagrams, which performed worse than similar text-only representations. Despite their advantage, iconic displays still only generated correct answers less than 50% of the time in Brase’s (2008) study. To further improve performance, HCI designers may need to consider more than choosing an iconic or spatial display.

Problems with Venn and Euler diagrams

Why some prior research would find that Euler type diagrams would be at a disadvantage (Sloman et al., 2003; Brase, 2008) is initially puzzling. Euler diagrams are thought to be equivalent to the notations or partial models people are hypothesized to use by mental models theory (Stenning & Oberlander, 1995). Euler diagrams’ specificity ought to compel people to generate the appropriate internal representation (mental model) depicted in the visual. These diagrams may be failing to control people’s attention, which is one way diagrams can help (Larkin & Simon, 1987). These diagrams may also be failing to limit the number of interacting elements that must be comprehended simultaneously and may be exceeding working memory. Venn and Euler diagram processing may be similar to understanding visual animations (such as an animated sequence of pulleys) in which performance degrades as more items are needed to be held in spatial working memory (Hegarty, 2000). Conditional probability problems also involve math computation. Math problems do involve the visuospatial scratchpad (Trbovich & LeFevre, 2003; Zago & Tzourio-Mazoyer, 2002; Houdé, & Tzourio-Mazoyer, 2003). Performing the mental calculations required to solve these problems could be adding additional spatial working memory load, thus increasing the complexity of the problem.

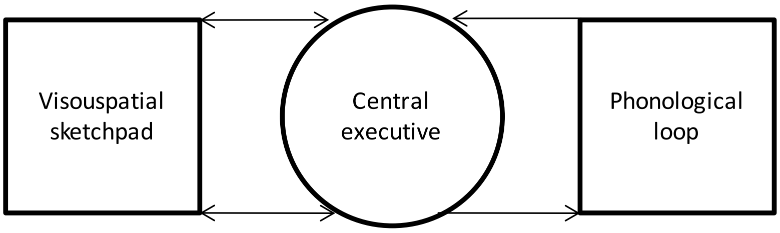
RELEVANT COGNITIVE THEORIES

Several cognitive models are relevant for this study, including theories of working memory, dual-coding theory, mental models, working memory capacity, graph comprehension theory, relational complexity and cognitive load. This study examines the relevancy for each of these theories and identifies relational complexity theory and cognitive load theory as critical theoretical frameworks that can guide visual construction.

Baddeley and Hitch's model of working memory

The model of working memory introduced by Baddeley and Hitch (1974) has become a dominant and well-researched model used in cognitive psychology, multimedia display and in some human-computer interaction (HCI) research. Working memory is normally defined as the involving dynamic and temporary storage and processing of information during the performance of complex cognitive activities (Miyake & Shah, 1999). Baddeley's (1986) model contains two subsystems: one for verbal information (the phonological loop, or verbal working memory) and one for visuospatial information (the visuospatial sketchpad, or visuospatial working memory).

The phonological loop (PL) maintains verbal information heard or read which decays in a few seconds unless maintained by a process known as articulatory rehearsal (Baddeley, 2003). The visuospatial sketchpad (VSSP) maintains visual information (object properties) and spatial information (or location/movement). Thus, the VSSP most likely contains two disassociated subsystems: a visual component, for processing object shape, color, texture and a spatial component for processing location, and motion (Klauer & Zhao, 2004). A third system, the central executive (CE), focuses the limited capacity in working memory and provides the capacity to switch or split attention across tasks (Baddeley, 2007). The three model system is depicted here:



Dual-coding theory

Most people, even trained physicians (Nelson et al., 2008), have problems with statistical reasoning indicating that they cannot retrieve a previously learned strategy (schema) from long term memory and apply it. Since few people have mastered Bayes' theorem, for most people, statistical reasoning and conditional probability problems appear to be novel problems. Dual-coding theory (Paivio, 1991, 1986) would predict that visuals can facilitate performance by aiding in retrieval of a schema from long term memory through referential encoding. According to dual-coding theory, schemas can be encoded in long-term both verbally and visually. A visual would provide an alternate means of retrieving from long-term memory the memorized schema for solving the conditional probability problem.

Since this problem is most likely novel and people do not have access to a relevant schema in long term memory, visuals, at least one designed with the intent of triggering references to schemas in long term memory, may not facilitate performance. Manipulations of the problem that include additional visual information (expressed verbally) may increase spurious referential processing. This could affect performance negatively (Knauff & May, 2006; Knauff & Johnson-Laird, 2002). Also, visual manipulations which bring irrelevant visual details to problem can impede performance, especially for low-working memory people who spend more time processing the irrelevant visual information than high working memory people (Sanchez & Wiley, 2006).

Mental models

Both behavioral and neuropsychological mental models theory research (Johnson, Laird et al., 1999; Ruff et al., 2003; Fangmeier, Knauff, Ruff & Sloutsky, 2006) shows that people reason in these problems not by following linguistic rules, but by means of visuospatial models of the problem structure. While verbal processing and the phonological loop would need to be employed to process the text of these problems and for certain math operations (Trbovivh & LeFevre, 2003), solving the problems requires manipulation of the premises and relations within the problem using visuospatial processing and central executive cognitive resources.

Dual task studies, which require participants to solve reasoning problems while simultaneously executing tasks that engage one of the components of working memory, help reveal which components of working memory are involved in a given task. Several dual task studies show that the VSSP and CE are involved in problem solving, consistent with mental models theory (Klauer, Stegmaier & Meiser, 1997; Knauf, Jola & Strube, 2001; Vandierendonck & De Vooght, 1997). Mental models need not be visual, and can represent non-spatial relationships, such as kinship, or non-visual precepts in deductive reasoning.

According to Ruff et al. (2003), model based reasoning interferes with other spatial tasks, not necessarily with concurrent visual tasks without a spatial component.

The Ruff et al. (2003) study is worth noting in that this study combines neurological data (data based on imaging of the human brain) with behavioral data (data based on users performance on a task). It found that reasoning was positively correlated with visuospatial tests (Block Design Test) and that reasoning (both the relational reasoning problems and the Block Design Test) activated regions in the brain for processing visual images and allocentric (object-centered) spatial processing as well as areas known for the preparation and execution of motor responses, all areas involved in visuospatial processing. This study shows that people with higher visuospatial skill as measured by the Block Design test also had confirming neurological data supporting additional visuospatial capacity. These participants also performed better on the reasoning tasks. People’s reasoning ability is limited by the capacity of the visuospatial system to manipulate multiple premises.

Normally thought to process visually presented information or visually imagined information, the spatial component of the VSSP is involved in reasoning (Ruff et al., 2003) and math problems (Trbovich & LeFevre, 2003; Zago & Tzourio-Mazoyer, 2002; Houdé, & Tzourio-Mazoyer, 2003). Conditional probability problems do have a math component, also requiring spatial working memory capacity.

Graph comprehension theory

Research into graph comprehension sheds some additional light on how visuals should help facilitate reasoning. Eye fixation studies showed that graph comprehension (e.g., line charts, bar charts) is an incremental, iterative process that involves encoding a visual pattern, interpreting the pattern and relating it to elements within the display. This process of recognition, interpretation and integration with different and more complex chunks repeats until a coherent interpretation emerges (Carpenter & Shah, 1998).

Cognitive processes can occur in a top-down fashion, guided by attention control mechanisms and goal-processing or in a bottom-up fashion driven by visual stimuli (Beck & Kastner, 2009; Knauff, 2009). Moreover, representation in the visual system is competitive with bottom-up and top-down cognitive processes influencing the neural competition (Beck & Kastner, 2009). When a visual object gains dominance in one cognitive system, it will tend to gain dominance in other cognitive systems (Beck & Kastner, 2009). Processing in perception starts in the retina and then proceeds to the occipital cortex, with follow-on processing by the dorsal ("where") and ventral ("what") streams in different parts of the brain. This bottom-up process responds to and is guided by perceptual processes that start with the eye. Visual processing can also work top-down, in which visual images in memory or guided by reasoning processes can evoke a pattern of neural activity that involve regions of the brain that overlap perceptual processing. Perceived and imagined visual objects and spatial relations use overlapping neural circuits (Knauff, 2009).

Viewers form a mental model of the quantitative information in the graph through an iterative process of identifying and relating visual patterns to associated variables (Shah, 2002). Through more automatic, bottom-up processing, visual features are mapped to interpretations and information not explicitly represented in the display is not easily inferred (Shah, 2002). The person's prior domain knowledge and graph reading skill interacts with top-down processing of a graph to affect interpretation. Skilled graph viewers or those with higher domain knowledge make better inferences from the data regardless of the graph format (Freedman & Shah, 2002). On the whole, even simpler graphs require a complex set of cognitive processes for people to comprehend them (Carpenter & Shah, 1998).

Cognitive load theory

Cognitive load theory (CLT) assumes a limited working memory system that can manage information for only a few seconds with nearly all information lost after about twenty seconds unless the information is kept active through a rehearsal mechanism (van Merriënboer & Sweller, 2010). These limits apply to novel information only, not to information already encoded in long-term memory. When dealing with long-term memory, working memory has no practical limitation (van Merriënboer & Sweller, 2010) and can process more information at one time. Working memory load can be affected in three ways:

- Intrinsic load, which is directly related to the nature of the task and is dependent on element interactivity which is defined as relational complexity
- Extraneous load, which is imposed by the task presentation procedures. How the information is presented will either reduce or increase extraneous load
- Germane load, which refers to working memory resources that can be allocated to managing intrinsic load. People will use germane processing to construct new schemas that represent the problem and link these new schemas to schemas already present in long term memory

The distinction between understanding and learning, while not often clear in CLT research, can be differentiated (Schnotz & Kürschner, 2007). Understanding occurs when all relevant elements of information are processed simultaneously in working memory. Learning occurs when changes in long-term memory occur. Learning can be explicit (learning is intentional and the person is aware and can verbalize what is learned) or implicit (person may be unaware of learning and may not be able to verbalize what was learned). Explicit learning requires increased germane load to further abstract, find patterns and link the information to other learned information (Schnotz & Kürschner, 2007).

While CLT is usually applied to learning environments, it has been used to examine problem solving and CLT researchers have found that techniques that facilitate partial understanding may fail to facilitate complete learning. Full comprehension of a complex problem may require some form of learning, explicit or implicit (Sweller, 1998, Pollock et al., 2002).

Working memory and short-term memory capacity

Working memory differs from short term memory. Cowan (2008) defines short term memory as “faculties of the human mind that can hold a limited amount of information in a very accessible state temporarily.” Miller, et al. (1960) refers to working memory as memory used to plan and carry out behavior. Cowan (2008) surmises that working memory is used to retain partial results while working out a math problem without paper or “to combine the premises in a lengthy rhetorical argument.” Baddeley’s (1986) influential model holds that verbal-phonological and visuospatial representations are stored separately and managed and manipulated by the central executive. In Baddeley’s model, working memory is generally viewed as several components (the phonological loop, the visuospatial sketchpad and the central executive) working together (Cowan, 2008). The model of working memory put forth by Cowan (2008) includes short-term memory within working memory..

The evidence for multiple buffers within working memory was supported by studies that showed verbal tasks interfered with verbal short-term memory but not visual short term memory and visual tasks interfered with visual short term memory and not verbal short-term memory (Brooks, 1968, den Heyer & Barrett, 1971; Baddeley, 1986; Baddeley & Hitch, 1974). Neuroimaging studies have shown that verbal short term memory relies primarily on left inferior frontal and right parietal cortices. Visuospatial short term memory relies on right posterior dorsal frontal and right parietal cortices and object-visual short term memory relies on left inferior frontal, left parietal and left inferior temporal cortices (Jonides, et al., 2008; Awh et al., 1996; Jonides et al. 1993). The neurological data supports the distinction of the verbal and visual short term memory stores with verbal short term memory with a marked left hemisphere dominance and visual-spatial and visual-object short term memory with a marked dorsal versus ventral stream split in the posterior cortices (Jonides, 2008, Ungerleider & Haxby, 1994; Baddeley, 2003).

Neurological research also shows the distinction between the central executive and the visual and verbal short term memory stores (Jonides et al. 2008). A meta-analysis of 60 functional neuroimaging studies reveals executive processing relies on dorsolateral frontal cortex and the posterior parietal cortex (Wagner & Smith, 2003).

Although early research into working memory capacity showed that people can recall about even chunks of information in short-term memory tasks (Miller, 1956), more recent and extensive research is placing a lower limit of about four chunks on the capacity limits in short-term memory (Cowan, 2000, Cowan, et al. 2005; Rouder et al., 2008) across both visual and verbal tasks. Cowan's extensive review (2001) identified the following studies that supported a smaller working memory capacity of about four items.

Table 1. Relevant studies showing working memory is limited to about four items

1. Imposing an information overload
    - 1.1. Visual whole report of spatial arrays (Sperling, 1960)
    - 1.2. Auditory whole report of spatiotemporal arrays (Darwin, et al. 1972)
    - 1.3. Whole report of unattended spoken lists (Cowan, et al. 1999)
2. Preventing long-term memory recoding, passive storage, and rehearsal
    - 2.1. Short-term, serial verbal retention with articulatory suppression (Pollack et al. 1959; Waugh & Norman, 1965)
    - 2.2. Short-term retention of unrehearsable material (Glanzer & Razel 1974; Jones et al. 1995; Simon 1974; Zhang & Simon 1985)
3. Examining performance discontinuities
    - 3.1. Errorless performance in immediate recall (Broadbent 1975)
    - 3.2. Enumeration reaction time (Mandler & Shebo 1982, Trick & Pylyshyn, 1993)
    - 3.3. Multi-object tracking (Pylyshyn, et al. 1994)
    - 3.4. Proactive interference in immediate memory (Halford, et al. 1988; Wicklegren, 1966)

4. Examining indirect effects of the limits
    4.1. Chunk size in immediate recall (Chase & Simon, 1973; Ericsson, 1985; Ericsson et al. 1980; Ryan 1969; Wickelgren, 1964)
    4.2. Cluster size in long-term recall (Broadbent 1975; Graesser & Mandler, 1978)
    4.3. Positional uncertainty in recall (Nairne, 1991)
    4.4. Analysis of the recency effect in recall (Watkins, 1974)
    4.5. Sequential effects in implicit learning and memory (Cleeremans & McClelland 1991; McKone, 1995)
    4.6. Influence of capacity on properties of visual search (Fisher, 1984)
    4.7. Influence of capacity on mental addition reaction time (Logan, 1988; Logan & Klapp, 1991)
    4.8. Mathematical modeling parameters (Halford et al., 1998; Kintsch & van Dijk, 1978; Raaijmakers & Shiffrin, 1981)

Cowan (2000) categorizes the literate he reviewed regarding working memory limitations into four groups (Table 1). In the first group, researchers have estimated capacity by using a large number of elements that overwhelm the person's ability to rehearse or otherwise encode the information. This approach lets them approximate the limits of working memory by measuring task performance with fewer or more elements to be processed in working memory. In the second category, researchers create experimental conditions that discourage rehearsal or recoding of new items using articulatory suppression (a secondary task in which the subject has to repeat an utterance during the presentation of items and perhaps during recall of the presented items), or present items that are too long to verbally rehearse. This track of research suppresses continual rehearsal as a technique for managing elements in working memory. Rehearsal may give participants an alternate strategy for managing larger numbers of elements in working memory that interferes with measuring working memory capacity. Suppression is a technique that helps researchers infer the limits of working memory capacity. The third group of studies includes research that shows a significant drop off in performance at about 4 items. People can manage a larger number of elements in working memory capacity via rehearsal or chunking. Some studies have noted that the time require to perform working memory tasks slowly increases from one to four items and then rises at a faster rate after four items. This performance discontinuity is another means if inferring working memory capacities, especially when non-working memory resources cannot be suppressed in the experimental task. The fourth category included research that, rather than relying on recall of presented items, observed capacity limits indirectly from the experimental task. In this category, Cowan includes Miller's (1956) seminal work on the magical number 7 +/- 2, and other studies showing that larger numbers of items are "chunked" together with each chunk containing 3-4 items aiding performance. While the object of these studies wasn't in attempting to directly test limits of working memory, the discovery of optimal chunk size is a means for inferring working memory capacity.

In Baddeley's comments on Cowan's review (Cowan, 2000), Baddeley welcomed the review of the lower number finding it persuasive and posited the lower limit of about four elements for worming memory capacity as being drawn from the central executive and its related storage system, the episodic buffer (see also Baddeley, 2010). In this open peer commentary portion of Cowan's review, of the 37 responses, two reviewers disputed his claim of "about 4" as a limit of working memory. The rest of the responses either strongly confirmed the lower limit of four, or acknowledged the limit but posed several challenges to Cowan's theory of a unitary store as being responsible for this lower limit. For the purposes of this study, the exact mechanism responsible for the lower limit of four items in working memory is not of concern. The insight that a large body of research is strongly indicating working memory is constrained to about four elements (or chunks) at a time may be useful for thinking about how to construct visuals to aid in conditional probability problems. Prior research into these problems has overlooked this key constraint.

When people solve problems, they rely on short-term memory of both the visual and verbal short term memory stores for maintenance of information needed for the problem. People also rely on limited central executive processing capacity to maintain focus and attention, and to manipulate the various components of working memory. This processing of the problem, while perhaps distinct from storage of information needed for the problem in visual or verbal form, is also limited to the relating of four entities (Halford, Wilson & Phillips, 1998; Andrews, Birney & Halford, 2006; Halford, Baker, McCredden & Bain, 2005; Halford, Cowen & Andrews, 2007; Awh, Barton & Vogel, 2007; Birney, Halford & Andrews, 2006; Saults & Cowen, 2007). Thus, both processing and storage are severely constrained to about four elements.

Relational complexity

Relational complexity is defined as the number of relations that must be processed in parallel, that is, at the same time, to perform a task (Halford et al., 1998; Halford et al., 2010). As relational complexity increases, the processing complexity of a cognitive process increases. While a cognitive process may be made up of several steps, the complexity of the process is the measure of the complexity of its most complex task. Process complexity is affected by how many interdependent elements have to be processed in parallel at one time, not how many have to be processed over time.

One example of a simpler relation is the “The dog is bigger than the cat” which, following Halford et al. (1998) translates to the binary relation bigger-than(dog, cat). Binary relations are similar to one-way experimental designs. Ternary relations contain three interacting elements are analogous to a two-way experiment design that contains a relation binding one dependent and two independent variables (Halford et al., 1998). The maximum relational complexity adults can normally process is the quaternary relationship which binds four interactive elements.

A simple task, such as selecting a restaurant based on how much money one has, can be described as a problem with a binary relationship (Halford, et al., 1998). The money-restaurant relation has a set of ordered pairs; in this case, an amount of money is associated with a restaurant (or a group of restaurants). The money-restaurant relation can be expanded to include importance as a factor. Important occasions might call for more expensive restaurants, although importance may have more influence when one has a lot of money than when one has little, in which case, the money factor has a greater influence (Halford, et al., 1998). Relational complexity is also analogous to the number of factors in an experiment design. Continuing with the Halford et al. (1998) explanation, a one-way experiment design is equivalent to a binary relation between the one independent and the one dependent variable (Halford, et al., 1998). Expanding the analogy, “a two-way experiment design is equivalent to a ternary relation between one dependent and two independent variables.” More interacting variables increases the number of observations or participants needed for a study, which is analogous to the increased processing load generated by problems with high relational complexity. Relational complexity is then, the number of interacting relations that are then bound to specific elements. Problems become more complex as the number of interacting factors increase. Processing complexity is related to the number of arguments in a relation (Halford, et al., 1998).

Halford et al (1998) note that complexity depends on the actual cognitive processes being employed and depends on individual attributes such as age, experience, and problem representation attributes. All of these factors will affect how task elements are chunked (collapsed into a single element) or segmented (sequenced rather than processed in parallel). Measures of processing complexity must be specific to the actual cognitive process used. Processing relations or premises often involves ambiguity than can be resolved when these relations or premises are considered jointly, or at the same time. Planning a correct strategy to solve a problem depends first on representing the complete structure of the problem. This planning process, if it has to process many relations, can impose high processing loads.

Relational theory predicts that people's need to process premises jointly will increase processing load (Halford, Wilson & Phillips, 1998).

The account of working memory put forth by Oberauer et al. (2007) has much in common with both a mental models account of reasoning and the relational complexity theory. Oberauer et al. (2007) says working memory capacity 'reflects the ability to keep several chunks of information simultaneously available for direct access' (p 50). Working memory is limited in the mind's ability to establish multiple and simultaneous temporary bindings. This limits a person's ability to construct new relational representations (p. 69). Mental model construction is constrained by the complexity of the model itself and the working memory capacity, especially the ability for the person to focus attention to the features of the model: Working memory capacity relies on:

> *"… a mechanism that quickly establishes and dissolves temporary bindings between these elements and positions in a cognitive coordinate system, or placeholders in a schema. Direct access to a multitude of separate elements is necessary to construct new relations between them and integrate into new structural representations. The limited capacity for relational integration is the most important limiting factor for reasoning ability."* (Oberauer, et al., 2007).

In a neurological study, Kroger, et al. (2002) confirmed that increasing relational complexity from 0 to 4 in a reasoning problem dramatically increased the percentage of errors and the time in processing, as well as increasing activation in areas of the brain associated with working memory.

Cognitive load theory and relational complexity in HCI Research

Both relational complexity theory and cognitive load theory appear as constructs in human computer interaction research. A body of research has examined what contributes to task complexity within the realm of safety and ergonomics, design of instructional interfaces and the correlation of physiological measurements with task complexity (Crosby, Iding & Chin, 2003). Researchers have applied relational complexity theory to examine complexity of interfaces for air traffic control devices (Xing 2004, 2007, 2008), to usability analysis for consumer products (Lo & Helander, 2004) and to computer security data visualization (Suo, Zhu & Owen, 2009).

Cognitive load theory has been used to analyze entity-relationship diagram comprehension (Masri, Parker & Gemino, 2008), to relate cognitive load to user satisfaction with e-Commerce applications when higher cognitive load resulted in less user satisfaction, Schmutz, Heinz, Métrailler & Opwis, 2009) and to examine its impact on Internet queries (Dennis, McArthur & Bruza, 1998; Joseph, Yukawa, Suthers & Harada, 2006). Despite the more extensive research programs that look at relational complexity theory and cognitive load in learning and cognitive science research tracks, only a handful of studies have examined the role of cognitive load in human-computer interaction (Schmutz et al., 2009).

SYNTHESIS AND KEY FINDINGS

Conditional probability problems, as normally presented both in real-life and in much of the prior research, usually require people to maintain between three and seven distinct elements or entities. While techniques for simplifying the process of solving conditional probability problems (e.g., natural frequency, clearly labeling and representing sets) have been researched, the role of visuals has been conflicting and very little of the prior research has focused on non-learning environments. In addition, none of the research has examined the impact the number of potentially interacting elements in the problem presentation may have on user performance. Further research into the relationship between the number of problem elements and the use of visuals to improve performance in real-world, every day problem solving is in order.

To improve performance, the display would need to take into account that conditional probability problems require people to build mental models in a limited working memory. Mental model premises and relations between elements are processed in visuospatial working memory, which varies in capacity from individual to individual and is limited. As the research into working memory capacity and relational complexity shows, this limit is most likely about four independent items processed at a time. According to graph comprehension theory, visuals can help by providing more automatic bottom-up processing that can guide or limit the focus of attention to a smaller number of appropriate elements of the problem. In addition, by shifting elements of the problem from a verbal form to a visual form may help people balance processing across multiple components of working memory (the phonological loop and the visuospatial scratchpad). By carefully controlling the display to address working memory constraints, including keeping the user's focus of attention on a limited number of elements of the problem at a time and use visual elements in lieu of text wherever reasonably possible, extraneous cognitive load introduced by the presentation of the problem is reduced, allowing the user to then effectively apply germane cognitive load.

CHAPTER 3. HYPOTHESIS DEVELOPMENT

Conditional probability problems typically require people to maintain four distinct elements or entities (the probability of event A and event B co-occurring, event A without event B, the probability of event B without A and the probability of neither event A and B occurring) or minimally three distinct elements (a sample population, a positive diagnostic population and a positive incident population). Depending on how the conditional probability is presented, it may contain up to six or seven elements, for which people have to select the correct three or four elements needed for a correct inference.

To select the proper elements, people have to read and process the problem text, build a mental model that depicts the relationships between each of the elements (in this case the nested set structure), refer to the appropriate numbers within the display (a frequency or a percentage), confirm the mental model comports to the external representation (the visual display), and then carry out the appropriate math to calculate a correct answer. Both the maintenance of each of the elements, (along with their labels and numbers) and the processing required to comprehend the set of relationships between the elements and validate among multiple competing relationships, can exceed the storage and processing limit of four independent elements at one time. In order for people to be aided by visuals in solving these problems, the information display will invariably need to limit potential element interactivity.

Following the review of literature above, conditional probability problems have the following attributes that can cause difficulty for people. First these problems appear to be novel problems. Because of lack of deep familiarity with Bayes' theorem, people are unable to apply Bayes' theorem or an equivalent strategy from long-term memory. Second these problems require use of a limited working memory capacity to build a mental model depicting the state of affairs of the problem. Third these problems have a number of interacting elements that can potentially exceed the working memory constraint of simultaneously processing about four independent elements.

HOW VISUALS CAN IMPROVE PERFORMANCE

According to Halford, et al. (1998), relational complexity is the number of interacting elements (variables) that must be represented in working memory in parallel to perform a task. Conditional probability problems typically contain three and can contain as many as six or seven elements that might need to be processed in parallel. Based on relational complexity theory, mental models theory, cognitive load theory and the theories of working memory discussed in Chapter 2, for visuals to reliably help with these kinds of problems, designers need to consider the following:

A. Through perceptual cueing mechanisms such as highlighting and controlling what is visible in the display, restrict the user's focus of attention to the task of relating two independent elements at a time. This should help people, especially those with low working memory capacity, by controlling the building of a partial mental model, thus facilitating the development of a full mental model.

B. Use visuals to depict relationships between entities rather than describing the relationship with words. In other words, replace words with images to reduce the load on working memory and the need for the user to switch between verbal/linguistic reading of text and visual/spatial diagram processing.

C. Repeat the process of cueing and processing fewer rather than more elements at a time to help the user make a series of correct inferences. Since working memory is conserved by preventing the number of independently interacting elements from exceeding four items, this should aid the user in building and validating a mental model and improve the accuracy of additional inferences.

Using this approach, the study expects to show that people's performance (speed and correct inferences) and satisfaction with the display will improve when the visual displays minimize element interactivity, encourage building partial models by limiting displays to fewer interacting elements at a time and substitute images for words to depict relationships. Prior research may have inadequately controlled the relationship between visual and textual elements as these guidelines suggest. If shown to be true and these display guidelines are followed, Euler and Venn type diagrams might be safely used in conditional probability problems.

HYPOTHESES

Based on relational complexity theory, diagrams with less element interactivity should perform better than diagrams with high element interactivity. Since working memory is limited to a small number of elements that can be independently related to each other (about four) (Halford, et al., 1998), diagrams that reduce the number of independently interacting elements will help users solve conditional probability problems better than diagrams with more independently related items. When element interacting interactivity is too high, working memory will be overloaded, negating the benefits of a diagrammatic representation. This study will compare the performance for Venn and Euler-like diagrams and text-only representations with low relational complexity (three interacting elements) and high relational complexity (six interacting elements).

Based on the relevant literature, for conditional probability problems the designer of the display must limit the user's attention to a few elements at a time. This can be done by giving the user perceptual cues such as shading, color or sequencing to focus their attention on a small number of elements at a time. Controlling the sequencing of processing the various elements reduces element interactivity. This approach should help people build a partial mental model and integrate the partial model into a full model where the correct inference can then be made (Pollock, Chandler & Sweller, 2002).

Neurological research into reasoning problems like the conditional probability problem shows central executive (CE) and visuospatial (VSSP) working memory involvement (Knauff, 2007, 2009; Kroger et al., 2002; Ruff, et al., 2003). The close relationship between CE and VSSP resources in reasoning is not surprising. Tests of spatial ability, especially the 2-D and 3-D manipulations within the paper folding tests and block design tests, may have a positive correlation with measures of central executive ability (Miyake, et al., 2001). For these reasons, the conditional probability problem involves both the VSSP and the CE for processing the visual (an external representation) and building the mental model to solve the problem (an internal representation). The display will need to reduce cognitive load so that the processing of the problem and the processing of the visual do not overload working memory resources. Keeping the potential element interactivity limited will save CE and VSSP working memory resources.

By using visuals to help solve the problem, designers can reduce the amount of text needed to depict the state of affairs. This eases the demands on the phonological loop, thus conserving working memory resources. By removing or reducing the text and using a visual display the user spends less effort processing the textual component to the problems. As users process the problem text, the user also starts to build a mental model of the relationships between the elements and may have to switch between the text and the visual. Visual elements that replace textual elements can offload some of this PL processing to the VSSP. This serves as a form of off-loaded memory to an external representation in which users can refer to items visually in the display, bringing elements back into working memory attention more easily (Larkin & Simon, 1997; Stenning & Oberlander, 1995). A visual display can also "constrain the range of possible cognitive actions" and thus prevent error and encourage proper conclusions (Zhang, 1997; Zhang & Norman, 1994).

Hypothesis 1

Under both high and low levels of relational complexity, users will demonstrate higher performance with diagrams than with text.

When the problem is simplified by controlling relational complexity (as in the low complexity treatments), demands on working memory will be reduced with either display (diagrams or text). Diagrams should help by further relieving working memory by substituting visuals in place of some text (reducing the amount of text processing needed) and by providing visual cues and guidance than can't be equivalently done in a simpler text-only representation. While the low relational complexity display, both in a text and a visual form, attempts to restrict the number of elements needed to integrate to three at a time, the visual form of the problem can provide additional visual cues that can guide users' attention using bottom-up perceptual processes.

Even if the problem has high relational complexity, diagrams should still facilitate performance. Since working memory is predicted to be overwhelmed in the complex, text-only display, visual aids can have an effect by substituting some of the text processing with diagrams and providing visual cues that can guide users' attention via bottom-up perceptual processes. This ought to help users determine the relevant sets to consider while determining an answer. While performance in the diagram or text-only condition is expected to be strongly influenced by complexity, the role of the visual under these two conditions is of interest. Visuals should be able to provide support for both low and high complexity versions of this problem.

Hypothesis 2

Under both high and low levels of relational complexity, users will have higher satisfaction with diagrams than with text.

Since the purpose of this study is to examine performance in a non-learning environment, it will be important that users perceive the displays as being easy to use and facilitating their performance. In of order for displays to be used in everyday, voluntary adoption contexts, user satisfaction will be important. Perceived usefulness and perceived ease of use (Davis, 1989) are two well-established constructs for predicting technology adoption. Davis (1989) defines perceived usefulness as “the degree to which a person believes that using a particular system would enhance his or her job performance.” Perceived ease of use refers to “the degree to which a person believes that using a particular system would be free from effort.” Together, these two constructs have explained nearly 40% of the variance in an individual’s intention to use technology (Venkatesh, Morris & Davis, 2003).

Hypothesis 3

Under both the text and diagram conditions, users will demonstrate higher performance with problems of lower relational complexity than with problems of higher relational complexity.

Since working memory demands will be reduced with low relational complexity displays that have less element interactivity, users will generate more correct answers. Some of the problems with the use of visuals in prior research may be a result of researchers failing to consider the number of interacting elements in conditional probability problems, the overall working memory demands of these problems and the limited working memory resources available.

Hypothesis 4

Under both the text and diagram conditions, users will have higher satisfaction with problems of lower relational complexity than with problems of higher relational complexity.

Relational complexity is likely to have an impact on satisfaction. A low relational complexity display will place less cognitive demands on the user and as a result will be likely to be perceived as easier to use. The perceived ease of use component of the adoption model proposed by Davis (1989) refers to the degree a person believes that using a particular system would be free from difficulty or great effort. In most real-world situations, adoption of particular displays by both sender and receiver of the communication is likely to be voluntary. As the TAM adoption framework (David, 1989) has shown, perceptions of ease of use and usefulness affect adoption.

CHAPTER 4. METHOD

This study recruited 158 participants from a large research university and examined the impact that text-only and diagrammatic representations with two levels of relational complexity have on participant performance and satisfaction with conditional probability problems.

PARTICIPANTS

This study used both students and staff participants at a major research university with varying ages and backgrounds in order to better approximate real-world populations. Many prior studies show participants' accuracy on conditional probability of problems is low to intermediate, ranging from 6% of the problems correct to 62% depending on the textual format (Brase,et al., 1998; Gigerenzer & Hoffrage, 1995). Most of these existing studies use top-tier national university undergraduate students as participants (Brase, Fiddick, and Harries, 2006) with greater working memory capacity than the normal population, which decreases over time (Hale, et al., 2008).

What this means is that a person in a real-world setting is likely to have less working memory capacity than a participants used in typical studies. Most real-world situations, such as a health care professional explaining the probability of having a disease given a positive test result or a risk management professional attempting to explain the probability of a sequence of events, are likely to involve people with a wider range of working memory. Interface designers will need to develop improved methods for conveying this information to people with a variety of working memory capacity using displays that can be accessed via the Internet (for example, electronic health care), or within computer-based decision support tools. By relying on a more diverse population, this study's results may have greater generalizability.

Participants were recruited across via email notifications and through regular recruitment methods used in colleges. Participants included students from several colleges on campus and staff and faculty members. About one-third of the participants were older than the traditional undergraduate college student.

EXPERIMENT DESIGN

This study used a 2X2 factorial design with the two factors being complexity and display type. Complexity has two levels: low and high. Display type has two levels: text-only and diagrams. This constitutes four treatments: simple text (3T), complex text (6T), simple diagram (3D) and complex diagram (6D).

Table 2. Experiment design

		Relational complexity	
		Low relational complexity	High relational complexity
Display type	Text	Simple text (3T)	Complex text (6T)
	Diagram	Simple diagram (3D)	Complex diagram (6D)

Because participants were randomly assigned to one of the four groups in the factorial design, this is a between-subjects design. High relational complexity problems contained six (6) interacting elements participants had to process before integrating them and computing an answer and low relational complexity problems had three (3) interacting elements to consider before computing an answer. Each problem asked a question in which the user has to infer a posterior probability. For each problem, the user had to mentally select among the different elements which ones need to be combined for a final number and determine the proper mathematical relationships between the elements. In order to restrict the problem to fewer interacting elements, low relational complexity diagrams let the user select and combine three or fewer elements at a time in two steps with the last step containing the question with the measured answer.

The independent variables are relational complexity (low relational complexity and high relational complexity) and display type (text-only and diagrams). The nature of the conditional probability problem is such that the four critical sets participants have to consider (H, not H, D and not D) can be expressed linguistically in two ways: with six interacting elements or three interacting elements (see the discussion below on treatments). This split also corresponds well with the working memory limit of about four independent elements. This convenient split in representation inherent in the conditional problem will allow this study to examine the role of relational complexity and display type on participant performance while considering working memory limits. The dependent variables will be performance and satisfaction (discussed in a later section).

TREATMENTS

Four treatments were used. Text only treatments with low and high relational complexity and diagram treatments of low and high relational complexity were used. For both text and diagram conditions, low relational complexity treatments had three interacting elements at a time and high relational complexity problems had six interacting elements at a time displayed. The diagram representation used Venn and Euler type diagrams to depict the relationship between elements of the conditional probability problems along with text to identify the visual elements and convey the problem background and questions. Text-only displays did not have any diagrammatic elements or visuals. A battery of ten conditional probability problems was used in each treatment, with some problems replicated from prior studies.

Applying Relational Complexity

This study used the number of potentially interacting elements in the problem representation as the measure representing complexity. In order to derive a metric representing complexity, the process model for the actual cognitive task needs to be considered (Halford, et al. 1998). Applying Shah, Mayer and Hegarty's (1999) description of the dominant model for graph comprehension to the problems considered in this study, the users of a display have to:

1. Read and process the problem text and identify the important elements
2. Build a mental model that depicts the relationships between each of the visual and/or textual elements (in this case the nested set structure)
3. Relate the elements to the appropriate numbers within the display and confirm the internal mental model comports to the external representation
4. Carry out the appropriate math to calculate a correct answer.

These steps are done incrementally and interactively as users work through a problem (Shah, et al., 1999). In the case of conditional probability problems, the complexity of this task isn't just the arity of the final computed relation (in this study all problems, text and visual, low or high complexity are, in the final correct computation, ternary relations), but in the process for testing and finally determining the relevant elements to include in a final computation. The maximum relational complexity of the conditional probability task is likely to be encountered in step two of the iterative four step process.

Figure 1 (the Blue Cab conditional probability problem, see Appendix G) has five critical elements in the display which need to be related to the element stated in the problem question (20 future accidents). This problem has within it many binary relationships, a larger number of ternary relationships and at least one quaternary relationship (the basic conditional probability contingency table). The total number of relations possible from these six elements (derived from the formula p = n!/(r!(n-r)!) for calculating permutations in which n is the number of elements in total, r is the number of elements selected and p is the total permutations for the given selection) is 63 relations or sets.

For example, expressing the three elements in Figure 2 (Cabs tested, Cabs identified as Blue and 50 future accidents) as elements A, B and C, generates the following set of seven possible relations, including three unary, three binary and one ternary relation to consider: {A, B, C, AB, AC, BC, ABC}. The user has to decide which one (or more) of these sets need to be considered as valid, what type of relation exists (subset, superset, conditional relationship) between the elements and what mathematical computations to perform for the given relations. In studies of errors that people (both adults and children) make in solving conditional probability problems, people consistently incorrectly choose the set of elements (the relation) and incorrectly choose the mathematical operators relating elements (Zhu & Gigerenzer, 2006; Gigerenzer & Hoffrage, 1995), with children consistently choosing fewer (typically two) elements when incorrectly solving the problem rather than more.

Problem representations that keep the number of elements needed to integrate to three or less have significantly fewer possible sets for users to consider. Consistent with the comments from Halford et al. (1998) that despite different representations a problem that are isomorphic when solved correctly -- as they are in this study -- the planning process (step 2 above) requires users to represent the problem as a whole. Displays with more potentially interactive elements and thus more combinations of relations to consider may result in higher processing loads and more user error. This study used the number of potentially interacting elements in the high and low complexity conditions as the measure that captures the increased processing load that higher relational complexity imposes. Since three (3) interacting element generates up to seven (7) combinations of relations and is far less than the

63 combinations of relations in the problem representations with six interacting elements, the processing load for the less complex problem representations should be less and user answers should be more frequently correct.

The combination of verbal and visual cues in each of the treatments needs to help guide the user into building the right mental model and the establishing the correct formula for relating the elements needed. While the hidden structure of the problem is simple, the process the user may go through to identify the hidden problem structure is complex and the user is likely to be overwhelmed with a plethora of relations ranging in relational complexity and in total count to consider iteratively.

This study proposes that conditional probability problems with six potentially interacting elements have more processing complexity than problems with three potentially interacting elements. Participants are likely to test more combinations of elements and potentially build mental models with greater relational complexity when presented with problems that have more potentially interacting elements than fewer. This will cause participants to make more errors; those with less working memory capacity are likely to make more errors than those participants with more working memory capacity. Using displays that simplify the problem solving process by reducing the number of potentially interacting elements and by using diagrams to provide visual cues for identifying the correct relations ought to improve user performance.

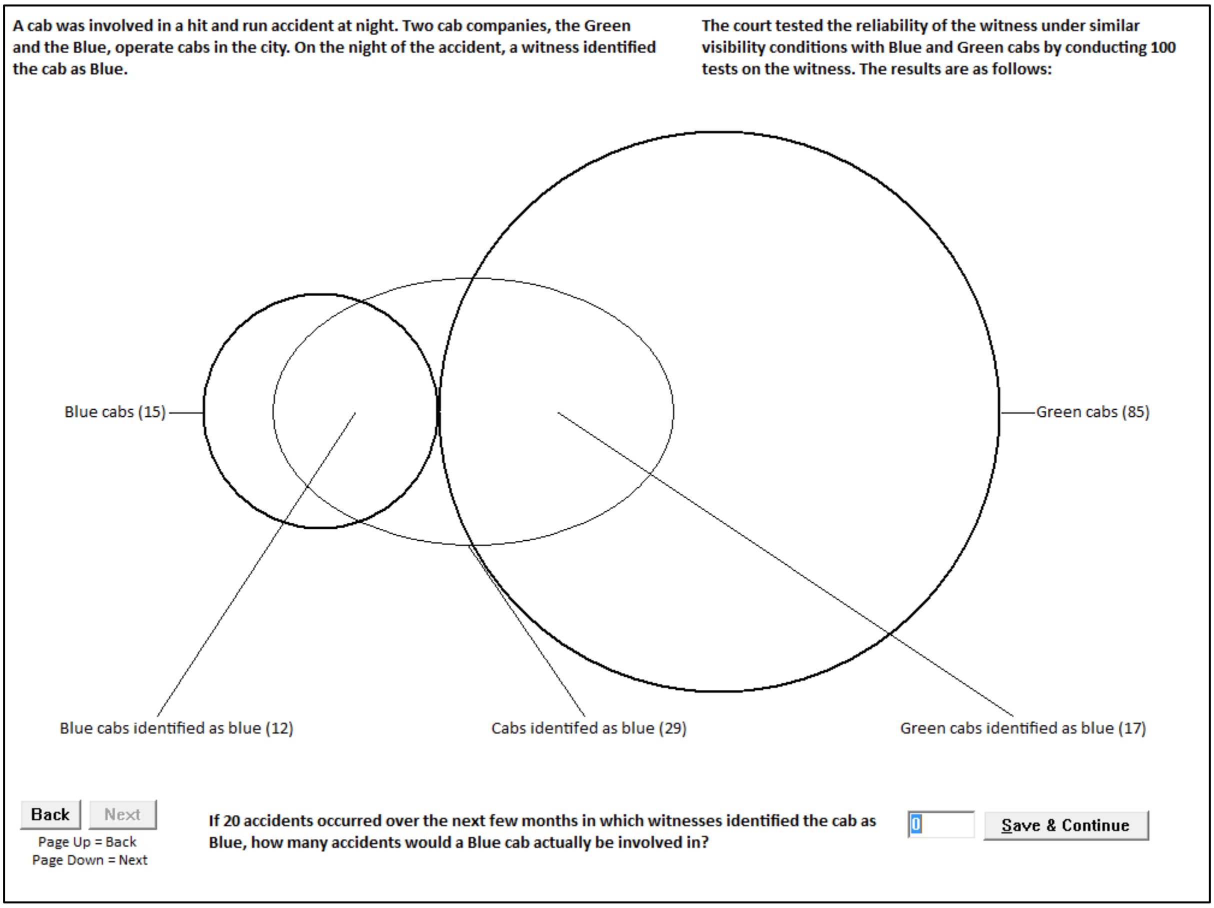
Diagram treatments

In prior research and in real life examples, depending on the presentation, conditional probability problems can have a high degree of element interactivity.

Diagram with high relational complexity (Treatment 6D)

Figure 1 shows this study's treatment of the conditional probability problem with a high number of interacting elements. This example has five separate elements, with two additional elements nested within one larger set (Cabs identified as blue). The problem asks users to examine the elements and identify which elements from the diagram ought to be included in a simple mathematical operation involving a sixth number contained within the problem question.

Figure 1. A diagram with a high relational complexity



In the above problem, the following elements are identified:

1. Blue cabs (15)
2. Green Cabs (85)
3. Cabs identified as blue (29)
4. Blue cabs identified as blue (12)
5. Green cabs identified as Blue (17)
6. Future accidents with cabs identified as Blue (20)

These six elements must be examined and a portion of them combined to compute the proper answer, in this case, the future accidents with Blue cabs identified actually being Blue cabs. The correct answer here involves the “Blue cabs identified as blue” set, the “Cabs identified as blue” set and the “20 accidents over the next few months” set. The correct mathematical computation is (12/29)*20 or 8.

The user’s challenge in this problem is to understand the problem text, maintain the linkage of the text labels to the visual elements, determine which elements are relevant to computing the answer, determine the nature of the relationship between the relevant elements and then to build a mental model (an internal representation inside the mind) that can then be used to produce a final answer. In this example, the mathematical model is in the form of (a/b)*c.

Diagram with low relational complexity (Treatment 3D)

An example of a conditional probability problem with fewer interacting elements is shown in Figure 2 and 3. In this problem, which in its final question is computationally identical to the more complex problem previously discussed, the diagram represents only two visual elements (and associated numbers) at a time and asks the user to integrate the problem text into a mathematical form for an intermediate answer (Figure 2).

Figure 2. Part one of a diagram with a low relational complexity

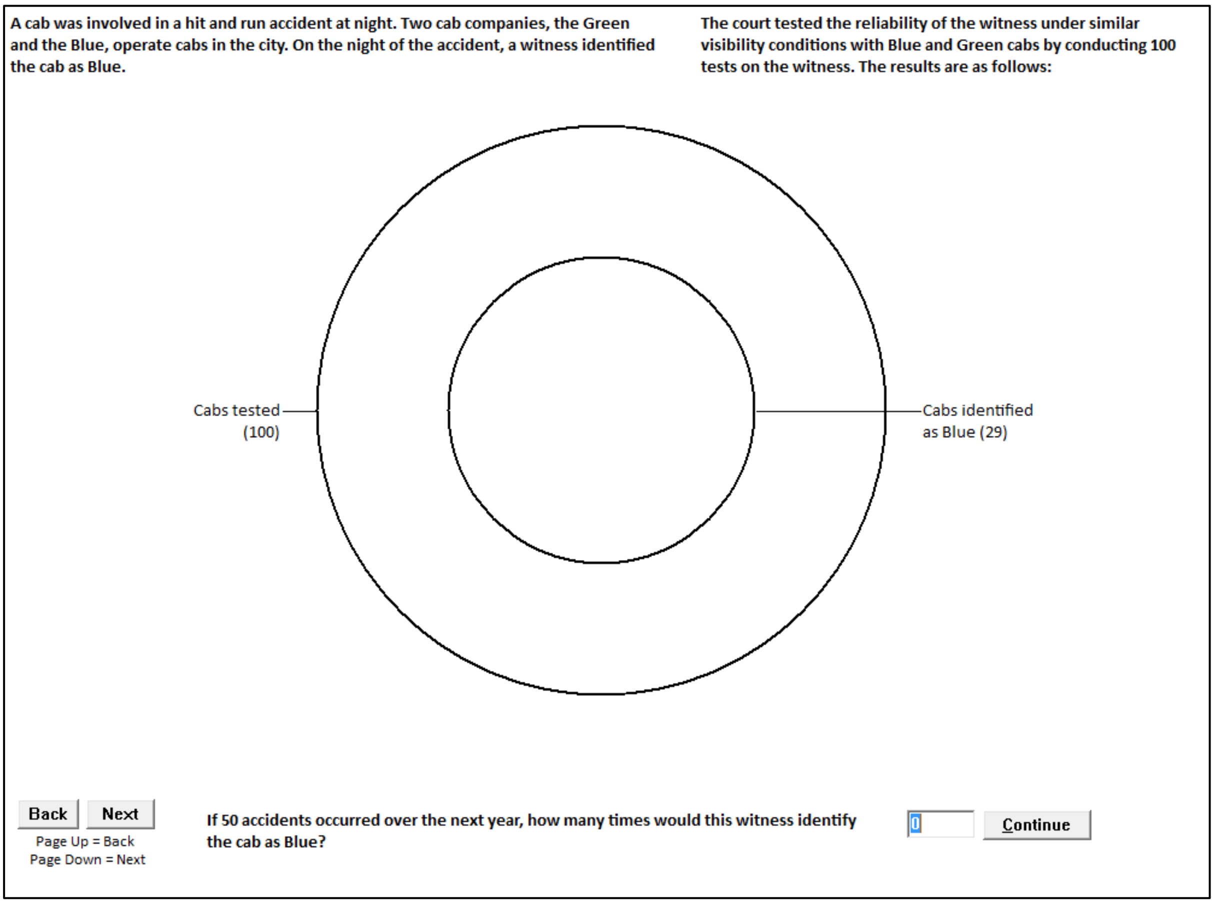
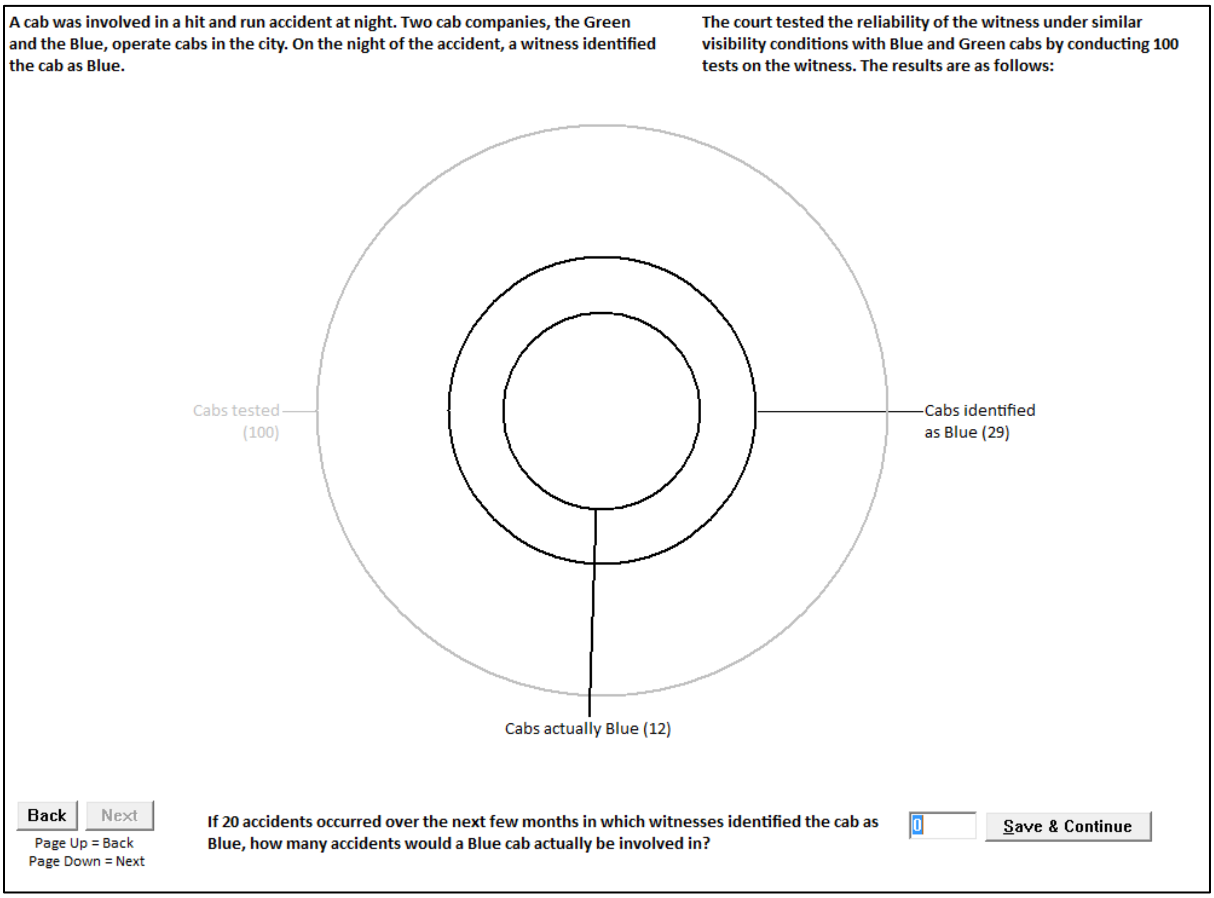


Figure 3. Part two of a diagram with a low relational complexity



In Figure 3, the problem text shows the user the final and complete visual, but explicitly greys out the initial element, indicating its lesser relevancy to the final problem question. The visual cue is designed to constrain cognitive action. By breaking the display into two steps and using perceptual cues to guide the user in lieu of text, relational complexity is significantly reduced. Using the visual to guide incremental mental model construction may help users shift needed elements in an out of attention, facilitating discovering and executing the correct computation. In the text of the above, simpler problem, the following elements are identified and clearly visible (not greyed out):

1. Cabs identified as Blue (29)
2. Cabs actually Blue (12)
3. Future accidents with cabs identified as Blue (20)

Again, these three elements must be examined and combined to compute the proper answer; in this case it is the number of future accidents with Blue cabs identified actually being Blue cabs. In this visual, the maximum number of elements to be included in a mathematical relationship is kept to three elements or fewer.

Text Treatments

For the text-only treatments, this study will use a similar approach, except no diagrams, shading or lines will be used.

Text display with high relational complexity (Treatment 6T)

High relational complexity problems will have six potentially interacting elements within the depiction of the problem which need to be related to the element indicated in the problem question. For example, the high relational complexity version of the Blue Cab problem is:

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. When the cabs were really Blue, the witness said they were Blue in 12 out of 15 tests. When the cabs were really Green, the witness said they were Blue in 17 out of 85 tests. Overall, the witness identified 29 cabs as Blue.
>
> Question: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

As in the diagram display type, this problem has the same six interacting elements, but without any diagrams or other visual elements.

Text display with low relational complexity (Treatment 3T)

The low complexity version of this treatment follows the same approach as the low-complexity diagram version (D3) first display requiring the participant to answer a question involving integration of just three elements at a time which is then followed with a second asking a question requiring the user to integrate an additional three elements. Unlike the diagrammatic representation, visual cues such as diagrams and shading will not be used.

> Display 1
> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. In 29 of those tests, the witness said the cab was Blue.
>
> Question: If 50 accidents occurred over the next year, how many times would this witness say the cab was Blue?
>
> Display 2
> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. In 29 of those tests, the witness said the cab was Blue. In the tests in which the witness said the cab was Blue, the cab was actually Blue 12 times.
>
> Question: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

Treatment Attributes To Control

In order to ensure generalizability and to avoid other confounding factors, various aspects of the conditional probability problem were consistently controlled across all four versions of the problem. The features controlled include: the use of natural frequencies; independence of complexity and visual representation, a consistent mathematical form; prevalence, sensitivity and specificity rates; clear labeling; use of diagrams, reference class size; distance effect; and comparability to other research.

Natural frequencies

Since the purpose of this study is to examine how reducing the number of interacting elements can facilitate people's performance with diagrams and since a large body of research shows that natural frequency representations elicit more accurate responses, the problems in this study all use natural frequency formats. Natural frequencies help because they simplify the math involved in calculating posterior probabilities and they make relationships between sets or elements clearer (Brase & Barby, 2006).

Independence of complexity and visual representation

In order to ensure that relational complexity is preserved in both the low and high complexity problems regardless of the display type (text or visual), the battery of ten problems maintain a consistent number of potentially interactive elements (3 and 6) in both the text and the visual representation. Because of this, the exact same number of possible relations to consider (7 and 63) is maintained in both the text-only and diagrammatic representation of the problems. Diagrammatic displays do use visual cues where possible (replacing visual elements in place of some words, using shading to guide focus), but preserve the same number of elements and possible relations to consider.

Consistent final mathematical form

Conditional probability problems follow a consistent mathematical formula based on Bayes theorem. As explained, natural frequency formats simplify the math involved. All these problem, whether they are medical problems or not, can be represented just like medical diagnostic information with base rates (the prevalence of a disease, e.g., 8 in 1,000 have cancer), sensitivity (e.g., for someone who has cancer, the test comes back positive 90% of the time) and specificity (e.g., for someone who does not have cancer, the test will be accurate 95% of the time) rates. All problems, regardless of their visual or text-only representation, involve the exact same mathematical formula to carry out the computation needed for a correct answer and will have a final mathematical form of (a/b)*c.

Prevalence, sensitivity, specificity, labeling, reference class size

To ensure generalizability across different forms of this problem likely to be encountered in many real-world situations, the treatments were constructed with a range of prevalence, sensitivity and specificity rates ranging from 1-90%, 5-93% and 33-99%.

People have difficulty with conditional probability problems because the problem text (and possibly the visual) does not make clear to what subset each of the probabilities belongs (Macchi, 2000). Therefore, in each treatment, all labels and text have been worded to clearly indicate how labels refer to visual elements and how the problem text refers to the diagram labels. Also, some research indicates people may make errors in answers based on reference class size (size of the populations within each problem). When people are asked to make judgments on smaller population sizes (less than 100), some biases disappear (Brase, 2002). Treatments in this study have a range of reference class sizes from small (less than 50) to large (in the thousands) to ensure generalizability.

Use of diagrams

The visual treatments will use Venn and Euler type diagrams. While some prior research into conditional probability problems has found these types of diagrams to be at a disadvantage in aiding users, Venn and Euler diagrams are widely used in everyday use and in diverse research problems. In order to prevent additional confounding factors from complicating this study and in order to give insight into the factors that can affect diagram performance, this study used a consistent diagrammatic representation for all visual problems. The use of diagrams allowed for a 10.4% reduction in word count compared with the text-only versions of the problems.

Distance effect

The relative size of the key classes in these treatments may matter. The distance effect, in which "discriminating two numbers that are numerically far apart is easier than discriminating two numbers that are numerically close" (van Opstal, et al., 2008) is well established in numerical cognition research. The set of problems in this study has a range of "distance" of the two key numbers (a/b) from 1 (where the two sets are identical in size) to 12 (where one set is 12 times larger than the other set). Problems in which the relative distance between the two key elements numbers is greater (e.g., 1 and 50 versus 34 and 45) may be easier for participants to distinguish and hence, easier for them to identify, hold separate in working memory and integrate into a mental model.

Use of problems from prior studies

Because the treatment manipulations within this study are designed to improve user performance, this study includes two classic Bayesian reasoning problems (the mammography problem and the Blue Cab problem), preserving the original prevalence, sensitivity and specificity rates for each problem. This will allow some comparability with prior research. Since much research has been done with text-only representations, the text-only displays can serve as a baseline group to provide comparability to prior research.

Other control variables and covariates

To test for individual differences and their effect on problem performance, participants completed a pre-test survey that collects background information (age, gender, educational level, probability reasoning experience, skill, and whether English is a native language). Because spatial ability may contribute to, if not be the primary source of user performance with regards to relational complexity (Halford, et al., 2007) and is used for processing spatial diagrams (Hegarty 2000, 1997), this study measured spatial ability and used it as a covariate. Based on prior research, spatial ability may interact with both complexity and display type since it serves as a predictor and a component of problem solving performance. Treating spatial ability as a covariate will let this study determine the impact the two main factors have on performance independent of spatial ability. The paper folding test from Ekstrom, French & Hardon (1976) was used to measure spatial ability.

DEPENDENT VARIABLES

This study has two dependent measures: user performance and user satisfaction. User performance was measured by one dependent variable: number of correct answers out of the ten in the battery of problems. User satisfaction was comprised of two confirmed factors from the Davis (1989) technology adoption model survey.

Participants expressed the answers to the conditional probability problems in whole numbers and were allowed to round up or down to the next integer. Since some participants may choose to do the math using percentages instead of fractions, correct answers are also allowed for different decimal places of percentages (e.g., 23%, 23.3% or 23.27%) when the percentage is used in an intermediate step in calculating the final answer.

User satisfaction was measured by a 14-item questionnaire based on Davis' (1989) technology acceptance model (TAM). In this model, perceived ease of use and perceived usefulness are two key determinants on user adoption of technology. Each item in the questionnaire was presented as a statement with a response on a 7-point Likert scale.

The hypotheses were tested by ANCOVA (for performance), ANOVA (for satisfaction) and t-tests comparing the means of dependent variables (number correct, perceived ease of use and perceived usefulness satisfaction scores) between the low and high complexity treatment groups and the text and diagram treatment groups.

PROCEDURE

The study was approved by DePaul University's institutional review board (IRB) while the lab sessions and data collection was completed at the University of Kentucky facilities. The University of Kentucky agreed to allow DePaul University be the IRB of record. Both parties (DePaul University and the University of Kentucky) signed an IRB authorization agreement authorizing the research to be conducted at the University of Kentucky with DePaul University as the controlling IRB. DePaul University approved the research project with exempt status. Two lab study supervisors and the principal investigator all obtained human research and ethics certification through the Collaborative Institute Training Initiative. The lab study supervisors managed the participant recruitment processed and helped manage session materials. Participants were recruited across campus via several approved means of approved recruitment including general email notifications and participation in the University of Kentucky's College of Communications research recruitment process. Participants were recruiting with an incentive lottery for two iPad tablet devices. The lab study supervisors administered the lottery process.

A small pilot session was conducted first to test the software and instruments and ensure the overall procedure worked correctly and determine the timing needed for the entire procedure. The study was conducted in a PC workstation lab setting in nine sessions over an eight week period. Participants had as long as needed complete the pre-test survey and the post-test survey and had 60 minutes to complete the 10 conditional probability problems. The spatial ability test was a timed test that took six minutes to complete and was delivered via pencil and paper rather than the computer. Participants first took the pre-test survey to collect non-identifying information (age, area of study, skill with probabilities, and level of education). This was followed by the paper folding spatial ability test, taken with a pencil and paper. After the spatial ability test, participants completed the suite of 10 conditional probability problems delivered via a computer. The pilot study determined that a maximum time of 60 minutes to complete the problems would be sufficient. Participants ended their session with the post-test survey which was also delivered via computer. The post-test survey contained 14 items based on the TAM (Davis, 1989) ease of use and perceived usefulness survey instrument and three questions related to perceptions of confidence. Appendix A contains example of the background information survey, the spatial ability test, the post-test survey of satisfaction. Appendix G contains details on each of the ten conditional probability problems. At no time did this study store or retain any personally identifiable code, marker or other personally identifiable data. Participants in the study were recruited with an incentive to possibly receive one of two Apple iPad devices that were chosen by lottery. A lab study supervisor performed recruiting and kept all information related to participants involved in the raffle separate from the study data and unavailable to the principal investigator.

While answering the suite of 10 problems, participants were allowed to use Microsoft Windows Calculator program to perform math calculations but were not able to use any other materials, devices or software. In order to ensure there were no biases resulting from the order of appearance for problems or elements, all problems were presented in random order. The high complexity diagram contains two primary Euler circles that were also presented in random order (left versus right circle first). Participants had 60 minutes to complete the 10 problems and all participants completed them within the allotted time.

CHAPTER 5. RESULTS

After concluding the data collection process, the data were evaluated to identify any errors in recording or software issues causing data quality. None were found and the software was reliable in recording data. The data was then prepared for further univariate and multivariate analysis.

DESCRIPTIVE STATISTICS

The data related to participants, performance and satisfaction were collected and analyzed. The cell counts, means, and standard deviations for the depended and covariate were calculated.

Participants

A total of 158 participants were included in the study. In order to have a more diverse set of participants, this study recruited students, staff and faculty from across the university community. Approximately one-third of the participants (57) were older than 22 and 28 of the participants were not in college. All participant characteristics are shown in Table 3.

Table 3. Participant demographics

Demographic characteristics		Count
Gender	Male	83
	Female	75
Language	Native English Speakers	143
	Non-Native English Speakers	15
Age	18-22 years old	101
	23-34 years old	30
	35-44 years old	8
	45-54 years old	10
	55-67	9
Highest educational background	High School	108
	Undergraduate	32
	Graduate	16
	Ph.D.	2
Degree in progress	None	28
	High school	7
	Undergraduate	110
	Graduate	7
	Ph.D.	6
Probability skills	1 – No skill	17
	2	28
	3	37
	4	41
	5	31
	6	4
	7 – Expert	0
Probability experience	Yes	105
	No	53
Visually impaired	Yes	16
	No	142

To test for the influence of participant background on performance, a regression analysis was completed on age, English as a second language, gender, self-reported experience and skill with probabilities, current educational level in progress and educational level attained and visual impairment. None of these variables were found to be significantly or marginally associated with performance (Appendix F). Participants were randomly assigned a treatment, resulting in an unbalanced, between-subjects design. The assignment of participants to treatments is listed in Table 4.

Table 4. Sample sizes

	Display Type		
Complexity	Diagram	Text	Total
Complex	40	37	77
Simple	44	37	81
Total	84	74	158

Spatial ability

Spatial ability was measured by the paper folding text (see Appendix B). This test has a minimum possible score of 0 and a maximum possible score of 20. Scoring for this test is the sum of all correct answers less the total of answers marked incorrect divided by four. For each marked (versus blank) incorrect answers, participants are deducted ¼ point. To support further analysis and visualization of the interaction effect, a post-hoc category representing spatial ability was created. A median split of the VZ-2 scores was used to divide the participants into two categories: high spatial ability (N=78) and low spatial ability (N=80). This blocking factor was used in post-hoc analysis. In this study, the participants' mean score was 10.24 and the median score was 10.5. Table 5 shows the descriptive statistics for spatial ability (VZ-2).

Table 5. Spatial ability score descriptive statistics

Statistic	Measure
Mean	10.24
Median	10.50
Standard deviation	4.19
Minimum score	0.25
Maximum score	20.00
High spatial average (n=78)	13.71
Low spatial average (n=80)	6.86
Shapiro-Wilk test for normality	W=.986, p=.12

Performance descriptive statistics

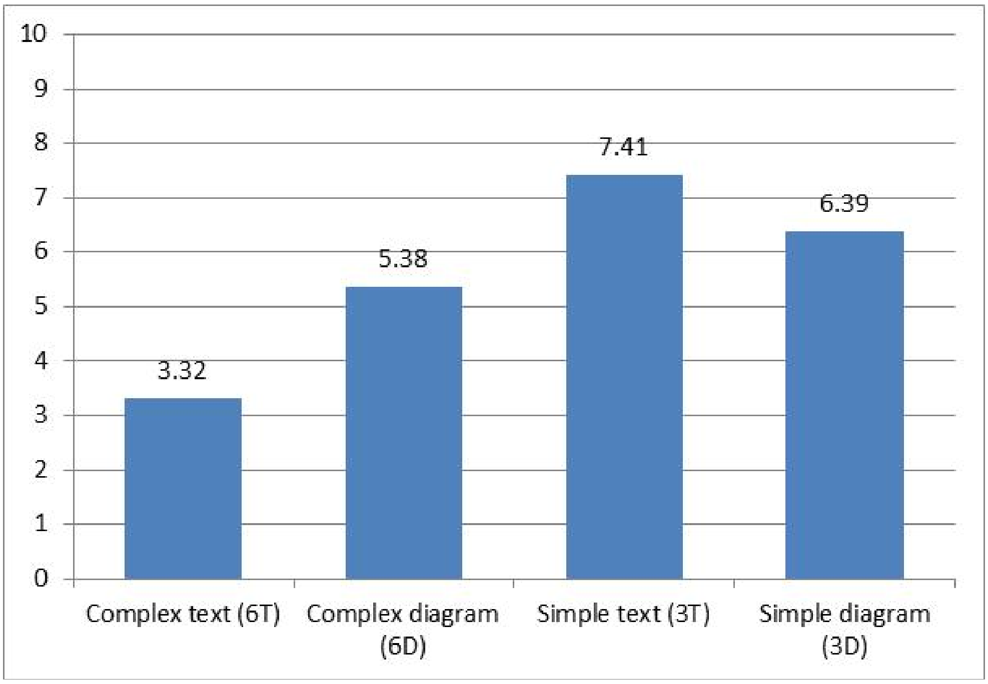
The means for performance and satisfaction (perceived ease of use and perceived usefulness) were calculated. For performance, the simple text display had the highest mean of

7.41, followed by the simple diagram display (6.39), complex diagram display (5.38) and complex text with the lowest mean score (3.32). Table 6 lists the means and standard deviations for performance for each display type. Figure 4 depicts these means.

Table 6. Correct means

Complexity	Display Type			
	Diagram		*Text*	
	M	SD	M	SD
Complex	5.38	2.62	3.32	2.00
Simple	6.39	3.60	7.41	3.13

Figure 4. Means for Correct.



Satisfaction descriptive statistics

User satisfaction was measured by a post-test survey derived from the Davis (1989) technology acceptance model (TAM). Perceived ease of use and perceived usefulness are two determinants of users' intentions to adopt technology. In the survey, seven questions measured perceived ease of use and six questions measured perceived usefulness. To establish discriminant and convergent validity of the two determinants, a rotated factor analysis was performed. One question from the perceived ease of construct and two questions from the perceived usefulness construct were dropped due to low factor loadings (see Appendix E). Cronbach's α values were calculated to check the reliability of the instrument and constructs. Perceived ease of use and perceived usefulness had high Cronbach's α values (.90 and .92 respectively) indicating the survey was reliable. The means for perceived ease of use and perceived usefulness are shown in Tables 7 and 8 and in Figure 5.

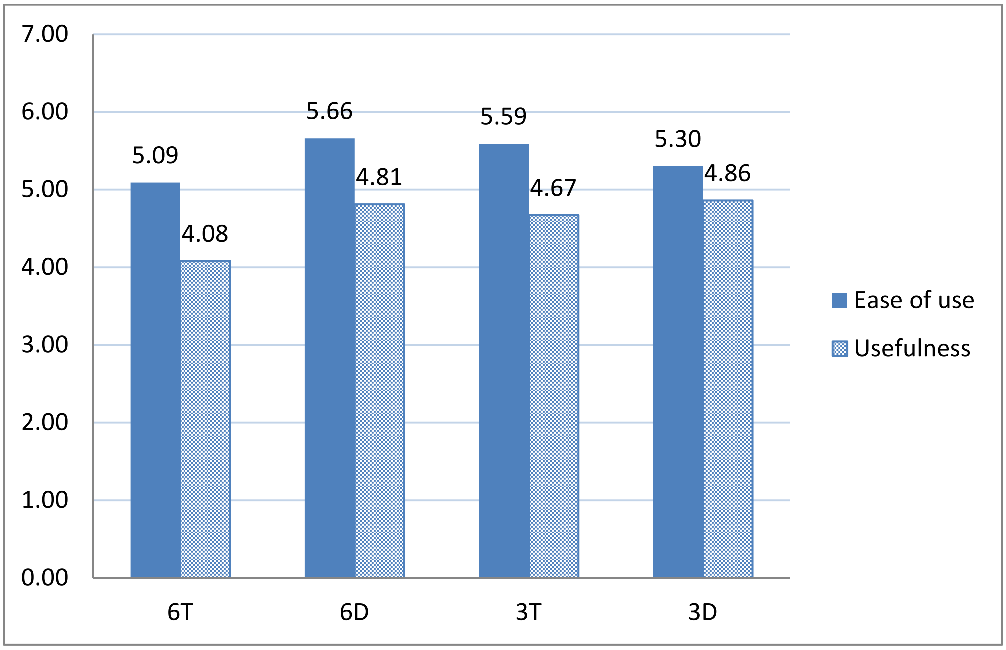
Table 7. Perceived ease of use means

	Display Type			
Complexity	*Diagram*		*Text*	
	M	SD	M	SD
Complex	5.66	1.21	5.09	1.30
Simple	5.30	1.47	5.59	1.35

Table 8. Perceived usefulness means

	Display Type			
Complexity	*Diagram*		*Text*	
	M	SD	M	SD
Complex	4.81	1.42	4.08	1.29
Simple	4.86	1.53	4.67	1.36

Figure 5. Means for satisfaction



HYPOTHESIS TESTS REQUIRED

To test the hypotheses for this study, a series of ANCOVA/ANOVA analyses were done followed by a series of one-tailed pairwise comparisons between displays. Since spatial ability is expected to impact performance, an ANCOVA analysis was conducted to analyze performance. For satisfaction, an ANOVA analysis was conducted for perceived ease of use and perceived usefulness. An ANOVA was performed for satisfaction because the literature review did not indicate an interaction between spatial ability and satisfaction nor was one hypothesized or planned for this study.

Performance analysis of covariance

Since the hypotheses called for the examination of two main effects (complexity and for display type), a two-way ANCOVA was conducted. The results of that ANCOVA are shown in Table 9.

Table 9. ANCOVA for Performance

Source	DF	SS	MS	F	Pr > F
Model	6	680.77	113.46	17.46	<.0001
Error	151	981.09	6.50		
Corrected total	157	1661.85			
Display type	1	11.47	11.47	1.76	.1860
Complexity	1	236.97	236.97	36.47	<.0001
Spatial_ability (covariate)	1	276.10	276.10	42.50	<.0001
Complexity*Display type	1	62.53	62.53	9.62	.0023
Display type*Spatial ability	1	75.93	75.93	11.60	.0008
Complexity*Spatial ability	1	18.30	18.30	2.82	.0953

The overall analysis of covariance showed that complexity and spatial ability affected performance ($F(6,151) = 17.46$, $p<.0001$, $\eta^2=.41$), but also identified a significant interaction between complexity and display type ($F(1)=9.62$, $p=.0023$), and between spatial ability and display_type ($F(1) = 11.60$, $p=.0008$). The display type main effect was not a significant effect in the ANCOVA analysis, nor was the interaction between complexity and spatial ability. For display type, overall participants did not perform better with diagrams than with text displays ($M=5.91$ versus $M=5.36$, $t(151)=1.30$, $p=.1970$). For complexity, overall participants performed better with simpler diagrams than with complex diagrams ($M=6.95$ versus $M=4.39$, $t(151)=6.23$, $p<=.0001$, Cohen's d=1.01). Details on the interaction between spatial ability and display type and complexity and display type are discussed further in this chapter.

Satisfaction analysis of variance

The ANOVA for perceived usefulness and perceived ease of use are shown in Tables 10 and 11. The overall ANOVA models for both variables were not significant. No differences between the treatment means for perceived ease of use and perceived usefulness could be identified. Users did not perceive diagram displays as significantly easier to use or significantly more useful than text displays. Likewise, users did not significantly perceive simple displays as easier to use or more useful than complex displays.

Table 10. Perceived ease of use ANOVA results

Source	DF	SS	MS	F	Pr > F
Model	3	7.89	2.63	1.46	.2264
Error	154	276.68	1.80		
Corrected total	157	284.58			
Display type	1	0.73	0.73	0.41	.5421
Complexity	1	0.23	0.23	0.13	.7223
Complexity*Display type	1	7.15	7.15	3.98	.0478

Table 11. Perceived usefulness ANOVA results

Source	DF	SS	MS	F	Pr > F
Model	3	15.02	5.01	2.45	.0604
Error	154	306.41	1.99		
Corrected total	157	321.42			
Display type	1	8.37	8.37	4.21	.0419
Complexity	1	4.07	4.07	2.05	.1546
Complexity*Display type	1	2.92	2.91	1.47	.2273

HYPOTHESES SUMMARY

The test of the four hypotheses could require up to 12 one-tailed comparisons. Hypothesis 1, that diagrams are superior to text displays for performance, requires two one-tailed t-tests between participants. The tests need to compare performance differences between simple diagram and simple text displays (3D>3T) and between the complex diagram and complex text displays (6D>6T). Looking at the first hypothesis, H1, participants in this study demonstrated higher levels of performance with diagrams than with text for only the high relational complexity (complex) treatments (5.38 versus 3.32 correct, $t(75)=3.84$, $p=.0002$, Cohen's $d=.87$). Participants did not demonstrated better performance with diagrams than with text in the low relational complexity (simple) condition (6.39 versus 7.41 correct, $t(79)=-1.35$, $p=.9090$). Thus H1 is partially confirmed.

Hypothesis 2, that diagrams are superior to text displays for satisfaction, requires four one-tailed t-tests. Satisfaction is comprised of two factors, perceived ease of use and perceived usefulness. Both variables representing these factors require comparisons across both levels of complexity. This results in the same comparisons as required for performance, but for each of the two satisfaction sub-factors: simple diagrams are superior to simple text displays (3D>3T) and complex diagrams are superior to complex text displays (6D>6T). Since the ANOVA for satisfaction and ease of use failed to detect significant differences, pairwise t-tests were not conducted. H2 is not supported.

Hypothesis 3, that low relational complexity displays are superior to high relational complexity displays, requires two one-tailed t-tests between participants. The tests need to compare performance differences between simple diagram and complex diagram displays (3D>6D) and between the simple text and complex text displays (6D>6T). Looking at the third hypothesis (H3), participants demonstrated better performance with low relational complexity (simple) displays over high relational complexity (complex) displays for only the text treatments (7.41 versus 3.32 correct, t(72)=6.68, p<.0001, Cohen's *d*=1.55.) Participants did not demonstrate improved performance with simple displays than with complex for the diagram treatments (6.39 versus 5.38 correct, t(82)=1.46, p=.0712). Thus H3 is partially confirmed.

Hypothesis 4, that diagrams are superior to text displays for satisfaction, requires four one-tailed t-tests. Satisfaction is comprised of two factors, perceived ease of use and perceived usefulness. Both variables representing these factors require comparisons across both types of displays. This results in the same comparisons as required for performance: simple diagram displays are superior to complex diagram displays (3D>6D) and simple text displays are superior to complex text displays (3T>6T). Since the ANOVA for satisfaction and ease of use failed to detect significant differences, pairwise t-tests were not conducted. H4 is not supported.

In reviewing the original hypotheses, this study partially confirms the hypotheses for performance but does not confirm the hypotheses related to satisfaction. Table 12 lists the performance hypotheses, the variable tested, the means, the resulting p test, and the effect size as measured by the Cohen's *d* statistic.

Table 12. Hypotheses t-tests for performance

	M	SD	M	SD	*t*-test	Cohen's *d*
H1 - Diagrams > Text: performance						
Correct						
3D>3T	6.39	3.60	7.41	3.13	-1.35	
6D>6T	5.38	2.62	3.32	2.00	3.84 ***	.89
H3: Low RC > High RC: performance						
Correct						
3D>6D	6.39	3.60	5.38	2.62	1.48	
3T>6T	7.41	3.13	3.32	2.00	6.68 ***	1.55

***$p \leq .001$. All tests are one-tailed.

H1 was partially supported. In the complex condition, diagrams improved performance compared with text. In the simple condition, diagrams did not help. H3 was partially supported. Simple text displays performed better than complex text displays. Simple diagrams did not perform better than simple text. Table 13 summarizes the hypothesis findings in data that partially support the four hypotheses.

Table 13. Hypotheses summary

	Measure	Test	Conclusion	Conditions	
H1:	Performance	Diagram>Text	Partially supported	High RC: yes	Low RC: no
H2:	Satisfaction	Diagram>Text	Not supported	High RC: no	Low RC: no
H3:	Performance	Low RC > High RC	Partially supported	Text: yes	Diagram: no
H4:	Satisfaction	Low RC> High RC	Not supported	Text: no	Diagram: no

The summary of the hypotheses t-tests (Table 12) and the depiction of performance means for performance (Figure 3) indicate that the less complex diagram display (3D) is failing to improve performance sufficiently. However, the ANCOVA results show a significant interaction effect between spatial ability (VZ-2) and the diagram display. Additional post-hoc analysis was conducted to understand the nature of this interaction.

SPATIAL ABILITY INTERACTION WITH DIAGRAM DISPLAYS

To understand the effect of spatial ability on diagrams, it is helpful to see a comparison of means by spatial ability. A median split of the spatial ability scores as a post-hoc classification illustrates the interaction (Figures 7 and 8). Figure 4 also approximates the plot of the slopes of the regression lines for each treatment in the ANCOVA analysis for performance.

Figure 6. Comparison of low- and high-spatial performance

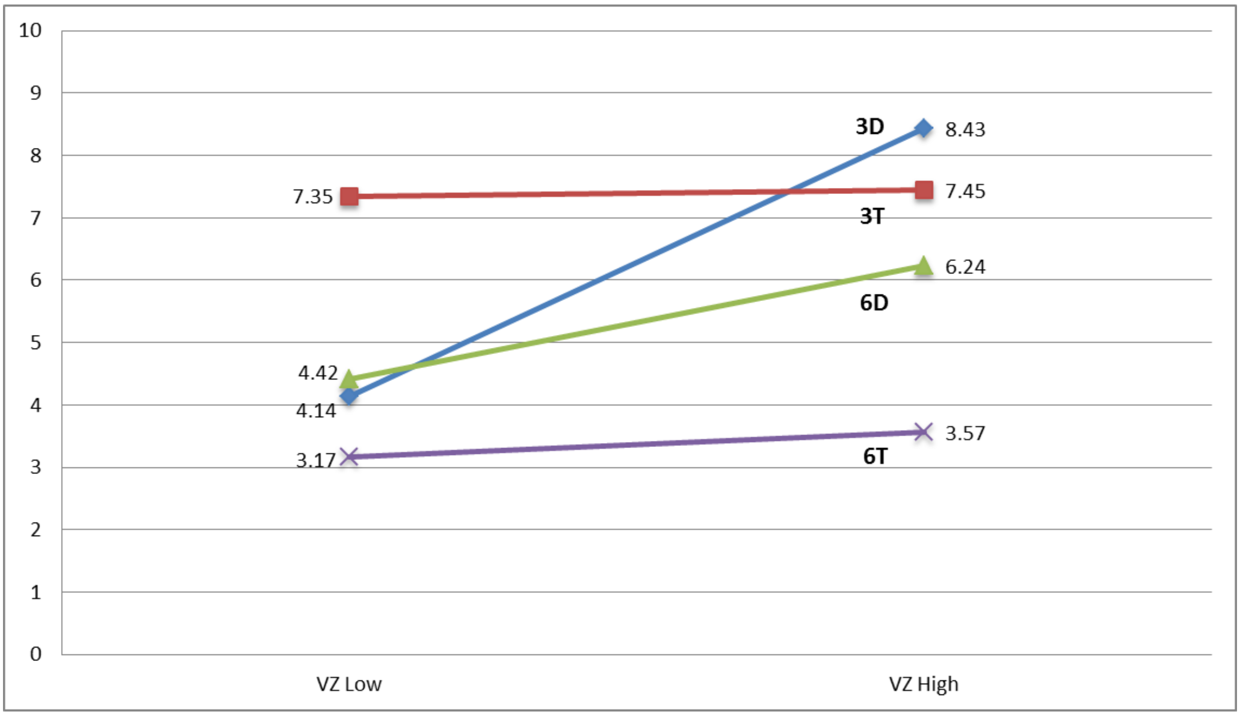
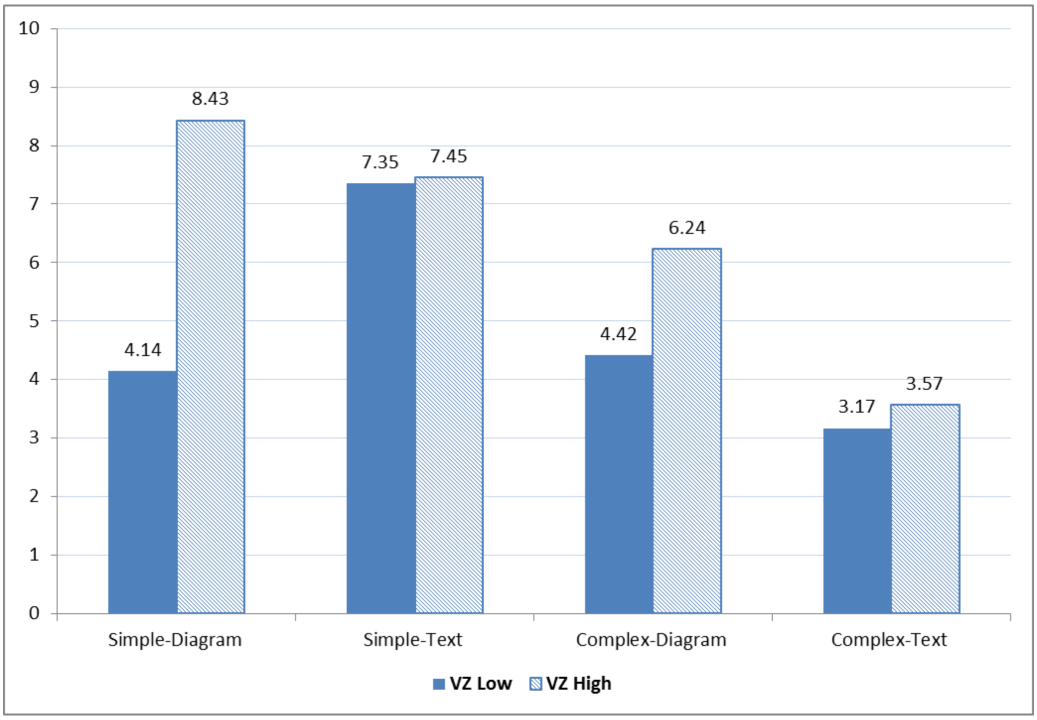


Figure 7. Comparison of low- and high-spatial performance



From Figures and 6 and 7, the diminished performance of the simple and complex diagram for low-spatial users becomes noticeable. Using the median split of spatial ability (VZ-2) as a method of differentiating high- and low-spatial users, two-tailed pairwise t-testing was conducted to verify the significance of the difference between high- and low-spatial user performance for all displays. While performing a median split can reduce the overall power of the analysis by as much as 38-60% (Cohen, 1983), the effect size of spatial ability is large and additional results presented here can be considered very conservative.

Low spatial participants performed 51% worse (4.14 versus 8.43, N=21 and 23 respectively) on the simple diagram and 28% worse (4.44 versus 6.24, N=19 and 21 respectively) on the complex diagram than the high spatial participants. Two-tailed t-tests confirmed this finding for both the simple diagram (3D) displays ($t(42)=4.79$, $p<.0001$, Cohen's $d=1.46$) and the complex diagram (6D) displays ($t(38)=2.31$, $p<.0264$, Cohen's $d=.74$). Also of note, low-spatial and high spatial participants did not differ in their performance on either the simple or the complex text-only versions of the problem ($p=.5662$ and $p=.9267$ respectively). The diagram displays significantly impaired low-spatial participants' performance. The text displays performed equally well for high- and low-spatial participants. Table 14 summarizes the high- and low spatial comparisons of performance.

Table 14. Spatial ability split comparisons by treatment

	High spatial		Low spatial			
	M	SD	M	SD	*t*-test	Cohen's *d*
Simple diagram	8.43	2.11	4.14	3.58	4.79 ***	1.46
Simple text	7.45	2.99	7.35	3.37	0.09	
Complex diagram	6.24	2.51	4.42	2.46	2.31 *	.74
Complex text	3.57	1.99	3.17	2.04	0.58	

$^*p \leq .05$ $^{***}p \leq .001$

Low-spatial user performance was diminished with diagram displays, especially the simple diagram display. This contributed to the interaction effect between display type and complexity. Diagram displays significantly impeded low-spatial performance. High spatial users can process diagrams for conditional probability problems in non-learning contexts.

CHAPTER 6. DISCUSSION AND IMPLICATIONS

Across the collection of ten conditional probability problems used in this study, people do better with diagrams when the problem is more complex and do better with few interacting elements in the problem when it is presented in a text-only display. Since this study used a collection of conditional probability problems that varied different problem features (prevalence, specificity, sensitivity rates, class size, and number of words), these findings are more generalizable than other studies that relied on a single problem (e.g., Brase, 2008). This study's findings also affirm relational complexity theory and mental model theory as promising approaches for guiding the design of visual displays for improving user performance in conditional probability problems. For high-spatial users, this study's findings also provide additional evidence that theories of working memory and graph comprehension can be applied to the construction of displays for conditional probability problems. However, the observation of an interaction effect between spatial ability and diagram displays paints a more complicated picture. Because of this interaction effect, this study's performance hypotheses were only partially confirmed. In order to improve performance uniformly across all users, one has to consider individual differences in spatial ability. What could be argued as a display that would be the most helpful (the simple diagram display), was found to significantly impede performance for users with less spatial working memory.

DIAGRAMS AND SPATIAL ABILITY

The relationship between spatial ability and user processing of diagram displays complicates matters for designers of systems. Why is there a degraded performance for low-spatial users? The theoretical models discussed in Chapter 2 provide an explanation. Diagrams are more likely to prevent correct mental model construction for people with less spatial working memory. As theories of working memory (Miyake & Shah, 1999; Baddeley,

2010) and mental models (Knauff, 2009) indicate, since spatial working memory is used to process the diagram displays and build a mental model, low-spatial users were noticeably impaired in constructing the mental model while processing the diagram display.

Based on the literature reviewed, solving these problems requires people to construct a mental model of the problem in their mind. Mental models, being spatial in nature, require the use of spatial working memory (among other working memory components). For those with less spatial memory, the use of a diagram creates additional demands on spatial working memory to process the visual diagram, impairing performance. The findings from this study seem to indicate that spatial working memory underlies the visuospatial sketchpad and contributes to not only processing visual information, but mental model construction. In addition, spatial working memory contributes to math computation (Trbovich & LeFevre, 2003; Zago & Tzourio-Mazoyer, 2002; Houdé, & Tzourio-Mazoyer, 2003). All three of these tasks (visual processing, math computation and mental model construction) are present in the conditional probability problems used in this study. Here, Baddeley's (1999) and Logie's (2011) account of the inner scribe component of the VSSP, which handles spatial relationships for visual and nonvisual content, aligns with the mental models account (Knauff, 2009) of the spatial nature of reasoning problems. In addition to its contributions to processing relations between visual objects, this amodel nature of the inner scribe might be used to also help construct mental models in reasoning problems generally and specifically in the conditional probability problems used here.

For low-spatial users, the diagram displays could be creating a disruption in more limited spatial working memory needed for mental model construction. The diagram displays may be exposing this dual role for the inner scribe portion of the VSSP. Because spatial working memory is strongly associated with executive control (Hegarty & Waller, 2005; Miyaki et al., 2001) and neurological mental models data shows tight integration between central executive and spatial processing capabilities in the brain (Knauff, 2009), low-spatial user performance problems with diagrams could be related to difficulty with central executive control, not just spatial working memory capacity limitations. Both Logie (2011) and Baddeley (1999, 2012) acknowledge the inner scribe's role in providing spatial processing for visual and non-visual inputs. Knauf (2009) argues forcefully for the role of amodal spatial working memory served by the posterior parietal cortex in reasoning processes. No research identified to date has investigated the potential conflict possible within tasks that have external representations that require spatial visual processing (diagrams) and reasoning processes that require construction and validation of mental models using spatial working memory (inherent in conditional probability problems).

How can low-spatial users' spatial working memory get disrupted by a visual? Cognitive processes can occur in a top-down fashion, guided by attention control mechanisms and goal-processing or in a bottom-up fashion driven by visual stimuli (Beck & Kastner, 2009; Knauff, 2009). Moreover, representation in the visual system is competitive with bottom-up and top-down cognitive processes influencing the neural competition (Beck & Kastner, 2009). Processing in perception starts in the retina and then proceeds to the occipital cortex, with follow-on processing by the dorsal ("where") and ventral ("what") streams in different parts of the brain. This bottom-up process responds to and is guided by perceptual processes that start with the eye. Bottom-up processing occurs before top-down processing (Van der Stigchel, et al., 2009). Visual processing can also work top-down, in which visual images in memory or guided by reasoning processes can evoke a pattern of neural activity that involve regions of the brain that overlap perceptual processing. Both bottom-up and top-down processes interact with each other as visual processing occurs, with top-down modulation of attention occurring in proportion that bottom-up processes have not resolved the salience of visual objects (McCains & Kastner, 2011). In other words, to the degree that bottom-up processing cannot automatically clarify visual elements, top-down processing focuses attention on the necessary visual elements.

In this study, participants' mental model construction was most likely guided by goal-oriented, top-down, cognitive processes along with more automatic, perceptual bottom-up processes attending to the diagram display. These bottom-up processes can trigger formulation of incorrect mental models based on input from the retina and corresponding activity from the visual cortex to working memory. In this mode of presentation, the act of perceiving the diagram through more automatic, bottom-up cognitive processes could have severely impaired the low-spatial users' ability to build the appropriate mental model. Low-spatial users may be less able to inhibit or restrict bottom-up processing, thus causing visual diagram processing to conflict with spatial working memory needed to construct and validate a correct mental model.

Knauff's (2009) account is of interest. It contends, based on neuroimaging studies of reasoning, that mental model processing critically involves two brain regions, the posterior parietal cortex (PPC) and the dorsolateral prefrontal cortex (DLPFC). The former is for maintenance and handling of amodal spatial information (the mental model) and the latter is for controlling the inspection and manipulation of the model. The PPC is also part of the dorsal "where" processing pathway that processes visual information. The prefrontal cortex is strongly associated with central executive processing (Knauff, 2009). Following this account, if individual differences in central executive functioning are the cause of low-spatial user impairment in the simple diagram condition, it is likely that low-spatial user performance with text-only displays would also be impaired. However, since the diagram display is the source of the impairment for low-spatial users, not text-only displays, it is reasonable to infer that amodal spatial working memory is overburdened for low-spatial users. Perhaps bottom-up perceptual processes invite spatial working memory activity to process the Venn diagram, preventing appropriate top-down mental model maintenance and construction for low-spatial users. If bottom-up perceptual processes initiated by diagram processing are inviting a kind of dual-task interference with mental model construction, this would be similar to other studies that have identified conflicts between visual representations of perceived stimuli and mentally generated ones (Zimmer et al., 2010).

Since spatial ability is also associated with other intelligence measures (Hegarty & Waller, 2005; Miyaki et al., 2001), one might also expect that low-spatial participants would have reduced performance on the text-only version of the problem due to presumed deficits in other non-visual cognitive capacities. However, this study found no significant differences in performance between high- and low-spatial users on text-only displays. A possible explanation is that the text-only displays facilitated problem solving reasonably well by using natural frequencies, freeing up spatial working memory similarly for both high and low-spatial users. While both high and low-spatial users scored worse on complex text-only problems compared with simple one, there were no significant differences between the high- and low-spatial groups for either the simple or complex text-only displays. Only when the visual was presented to the low spatial group was performance and presumably spatial working memory capacity impacted. Since this study relied exclusively on the computer presentation of each problem by not allowing any secondary tools such as notepaper to interact with, participants had no choice but to interact with the computer display to solve the problem. This study identified two sources of working memory constraints: 1) a capacity limit constraint predicted by relational complexity and 2) possible a controlled attention constraint induced by diagrams and especially simple diagrams that affects low-working memory users by inhibiting the appropriate top-down control over the reasoning process. While this study points to capacity limits in working memory as indicated by relational complexity theory as the cause of the variation in user performance, it also points to the interaction between bottom-up visual processes and top-down attentional control processes impairing low-spatial users when a diagram is used. Two related but different working mechanisms may be involved. An alternative explanation is that other tests of cognitive ability (such as need for cognition test or the cognitive reflection test) that are different from tests of spatial ability other related cognitive measures, may be better predictors of text-only performance (Toplak, et al., 2011).

Do the kinds of errors that users made support this explanation? In reviewing incorrect answers across all four treatments, it was possible to unambiguously categorize some, but not all, of the errors. For simple problems, participants can select up to four possible sets and for the complex problems participants can select up to seven. In addition, for the simple problems, participants can infer some of the ‘hidden’ sets by subtracting two sets from each other. For example, in the simple version of the Blue Cab problem, participants can easily infer that of the 100 cabs tested, for which 29 were identified as Blue, 71 were not identified as Blue. Participant can make predictable errors by using (or inferring) incorrect sets or relating them together incorrectly. Across all errors made, 53% could be classified as set identification and set relationship errors. These errors are consistent with the notion that people sometimes fail to properly identify the sets, infer the proper relationships between the sets and then carry out the mathematical operations consistent with the relationships. This indicates that the process of identifying and relating possible sets, which is part of the processes for constructing and verifying the mental model, may be undergoing degradation under complexity in general and under the diagram displays for low-spatial users.

Considering the three component model of working memory of Baddeley (2007) which includes a central executive, the visuospatial sketchpad and the phonological loop, there may be another plausible explanation. Low spatial users may be suffering from an attention distraction effect, especially with the simple condition. In this condition, an intervening question was placed in between presentation of the first two sets and the final set (Figure 2) in order to control relational complexity. This extra step may have been like a “speed bump.” In the simple diagram condition, low spatial users may not be able to effectively manage the coordination of verbal and visual working memory resources across the visual and verbal modes with this intervening question while simultaneously constructing a mental model. This represents a dual-task conflict. In this case the two tasks are processing the visual and building a mental model of the problem. Most designers do not normally think that presenting a simple diagrammatic visual with text introduces a dual-task conflict. It is quite possible that a similar impairment may be occurring in the use of Venn and Euler-type diagrams used more broadly across other related problems such as relational and syllogistic reasoning which may use working memory resources similarly (Knauff, 2009).

Low-spatial user performance with complex diagrams also deserves some discussion. Unlike the simple diagram condition, in the complex condition, low spatial participants’ performance did not reverse under the presence of a visual. Instead it failed to improve performance as much as it did for high spatial users. As been discussed elsewhere here, spatial working memory has to serve two tasks: process the visual and build the mental model. In the presence of a diagram (which consumes limited VSSP resources), low spatial users may have less spatial working memory capacity than high-spatial users to build the mental model and solve the problem. Automatic, bottom-up, visual processing induced by the computer diagrams may consume low-spatial users’ working memory needed for building the mental model. However, other working memory resources may be affecting performance. Does spatial working memory reliance on top-down attention mechanisms (the central executive) (Awh & Jonidas, 2001), cause a deficit in central executive coordination induced by the visual? Or is it a failure to maintain the visual image or the mental model (or both) in spatial working memory due to excessive decay in time-limited buffers (Cowan, 2001)? Future research is called for to clarify the mechanism underlying the performance degradation which can, in turn, give guidance to designers.

High spatial participants performed best with the simple diagram with a mean of 8.43 correct. While this measure was not significant in its difference from the next best performing treatment (simple text, mean=7.37) perhaps due to ceiling effects in which the most dominant score, the mode, was 10 (13 out of 37 cases), the score is quite high when compared with prior study scores of total correct (Brase, Cosmides & Tooby, 1998; Gigerenzer & Hoffrage, 1995). For high spatial users, simple diagram displays perform very well. The highest score in all conditions for the low-spatial users was the simple text condition with 7.35 correct. Again the most dominant answer and the mode for this group was 10 (13 out of 44 cases). These scores, which were across a battery of ten problems that varied in several aspects (content, sensitivity, specificity, prevalence, class size, and word length) should be fairly generalizable to other types of conditional probability problems. Following the theories of working memory and mental models discussed in this study, high-spatial users appear to have ample VSSP capacity to both process the visual and construct a mental model to infer a correct answer.

While other researchers exploring the use of visuals to enhance learning have noted a spatial ability effect where improved multimedia displays favor those with high spatial ability more than those with low (Mayer, 1997), the reasons for the improvement cited by Mayer (1997) are related to dual-coding theory where visual and verbal codes together provide for better recall than verbal codes alone. This study is a non-learning study and did not require participants to encode any schemas into long-term memory. Since participants brought little if anything from learned schemas such as the Bayesian algorithm for calculating conditional probabilities to the problem solving process, the interaction of spatial ability with diagram displays is most likely confined to working memory. Nonetheless, the cognitive theory of multimedia learning (CTML) guideline applies here as well. Diagrams benefit people with higher spatial ability and hurt those with less.

SATISFACTION AND POTENTIAL ADOPTION ISSUES

The small range in satisfaction means between the four displays were most likely due to the computer-generated displays that were designed to be relatively easy to use regardless of the problem complexity. In the design of the displays, care was taken to apply the known improvements for understandability, ease of computation and graphical display. This is supported by the mean satisfaction scores for both satisfaction measures which were above the midpoint of 4 which denotes neutral satisfaction. The means for the treatments ranged from 4.08 to 5.66 out of a seven-point scale.

The finding that satisfaction scores were similar across all displays begs a more difficult question. If users cannot perceive any differences in satisfaction between displays that hurt their performance from those that help, how will displays be used in situations of voluntary adoption, where users are free to choose whatever display they perceive to be best? If users' perceptions of satisfaction cannot guide them into the appropriate display, users may voluntarily, and unknowingly, choose a display that hurts their performance. For example, low-spatial users may be inclined to adopt simple diagram (3D) displays, unaware of the diagram's negative effect on their performance. Future research may help shed light on the relationship, if any, between satisfaction and adoption of computer displays for conditional probability problems.

COGNITIVE EFFORT AND PERFORMANCE

To further examine reasons why low-spatial users had difficulty with the simpler diagram display, this study examined two questions contained within the satisfaction survey that were relevant to perceptions of cognitive effort (Paas, 1992). These two questions were part of the TAM satisfaction survey (see Appendix B) and were asked of all 158 participants. These questions might be helpful in shedding light on whether or not low-spatial participants had more difficulty with the simple diagram (3D) display because of increased cognitive load. These two questions on the survey are relevant to perceptions of cognitive effort:

1. The comprehension task on this presentation required less mental effort.
2. I felt frustrated when performing the tasks for these problems.

These two questions were not included into the final two factors (perceived ease of use and perceived usefulness) that comprise satisfaction because of low loadings but did form a factor on their own. A full set of data was available for these two questions. All responses to these two questions were rated just as the other TAM questions were – on a seven point scale. The responses to these two questions were combined and analyzed to see if the low-spatial participants did report any increased difficulty with the diagram displays compared with their high-spatial counterparts. An post-hoc two-way ANOVA (with main effects of complexity and display type) of this measure was highly significant ($F(3,154)=15.38$, $p<.0001$, $\eta^2=.23$), with two highly significant effects for complexity and display type ($p<.0001$ and $p=.0012$ respectively) indicating the overall model is valid.

In the simple diagram condition (3D), low-spatial participants did report more difficulty than high spatial participants. Two-tailed tests confirmed the finding (4.43 versus 5.52, $t(42)=3.03$, $p=.0041$, Cohen's d=.92). However, in the complex diagram condition, low-spatial participants did not report more difficulty than high spatial participants (3.97 versus 4.74, $t(38)=1.81$, $p=.0781$, Cohen's d=.57). Low-spatial and high spatial participants were also indistinguishable in their reporting of difficulty with the simple text condition (4.74 versus 4.93, $t(37)=0.37$, $p=.7121$) and the complex text condition (3.22 versus 3.18, t(35)=0.10, p=.9232). The additional difficulty low-spatial participants expressed beyond what their high spatial counterparts expressed was reserved for the simple diagram displays only.

When combining both low spatial and high spatial participants together, participant perceptions of difficulty as measured with the two questions above differed as expected, depending on the overall treatment performance. Two-tailed tests confirmed that overall, participants rated complex diagrams as more difficult than simple diagrams (4.38 versus 5.00, t(82)=2.14, p=.0352, Cohen's *d*=.47) and complex text as more difficult than simple text (3.20 versus 4.85, t(72)=5.77,p<.0001, Cohen's *d*=1.36). Also, participants rated complex text as more difficult than complex diagrams (4.38 versus 3.20, t(75)=4.03, p=.0001, Cohen's *d*=.93). This confirms that users reported less cognitive effort for simple diagrams compared with complex ones and less cognitive effort for simple text compared with complex text. However, most likely due to the interaction effect between spatial ability and diagram displays, when pooled together, participants did not rate simple diagrams as easier than simple text.

Interestingly, while low-spatial users rated simple diagrams as more difficult than their high-spatial counterparts, neither the low-spatial nor high spatial groups differed among themselves in perceptions of difficulty between simple diagrams and simple text displays. Within each group, the difficulty of simple diagram and simple text displays were rated similarly. While between-group differences exist, none do within the groups. Low spatial users expressed more effort than their high spatial counterparts with the simple diagram display, but they did not report any difference in effort among themselves between the simple text and simple diagram displays. In other words, the low-spatial users were unaware of any increased cognitive effort for the low-spatial diagram. An analysis of the total time required to complete the problems revealed no significant differences between any of the displays, or between high- and low-spatial users for each display. If the time it took the user to complete problems is indicative of increased effort, users did not exhibit any differences in the total time they took.

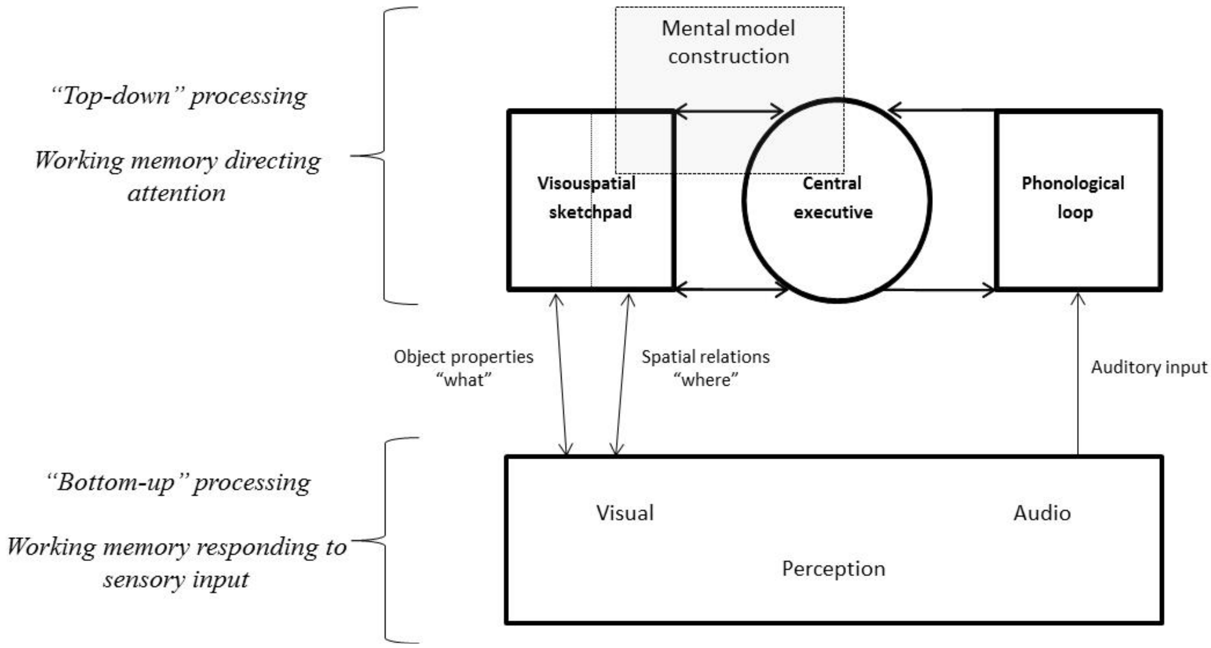
This has two implications. First, since perceptions of difficulty for low spatial users are the same for simple text and simple diagrams, rather than increasing cognitive load for those users, the simple diagram display may have *over-facilitated* the problem while still conflicting with spatial working memory, leading low-spatial users to easily select incorrect elements. This also suggests that users are largely unaware that bottom-up visual processes may be leading them astray. Working memory capacity may be related to central executive processing (Zimmer, et al., 2010). Individuals may differ in their ability to control attention and inhibit the processing irrelevant information (Zimmer et al., 2010, Oberauer et al., 2007, Feldman Barrett et al., 2004). Low spatial users may have had difficulty in effectively suppressing irrelevant information, especially in the presence of displays that facilitate performance like the 3D displays. While user performance was impaired in the simple diagram (3D) display, participant perceptions of effort did not detect the impairment. In the interaction between bottom-up and top-down cognitive processes, it might be possible that low- and high spatial users differed in their ability to suppress errant over-facilitation caused by more automatic bottom up processes. Low-working memory users may have a similar capacity for processing elements as do high working memory users, explaining their similar performance with text-only displays, but may have an inability to control attention in the presence of “seductive details” in visual displays (Sanchez & Wiley, 2006), and “trip up” in their answers. Low spatial users may benefit from a more complex diagram in order to prevent this over-facilitation. Again, future research is needed to clarify this point.

Second, this lack of a difference in perceptions of effort between displays within low- and high-spatial groups may have display adoption implications if cognitive effort proves to be a contributor to adoption of technology (Schmutz, et al., 2009) for these kinds of problems. If low- and high-spatial users perceive displays that compromise their performance similarly to displays that aid their performance, adoption of the wrong displays might be more likely to occur. While this study did not hypothesize effects of cognitive effort, future research may find perceptions of cognitive load might prove more useful and appropriate measures to predict adoption than satisfaction. While the two questions that comprise cognitive effort in this study are derived from a TAM satisfaction survey, they are similar to questions that describe cognitive effort. However, satisfaction and cognitive effort are different concepts. More robust measures of cognitive effort may be helpful in rating user perceptions of the displays for conditional probabilities that can be used to predict technology adoption. In this regard, future studies may help.

SUMMARY OF WORKING MEMORY IMPLICATIONS AND THE DIAGRAM DISPLAYS

Considering the initial three component model of working memory (Baddeley, 2012), users need to coordinate visual and verbal information in the diagram displays (Figure 8). The verbal information is held in the phonological loop and the visual information in the visuospatial sketchpad. The VSSP is comprised of two components that process object and spatial properties. Low-spatial working memory users may have less ability to suppress automatic bottom-up processing, especially if the visual image has high saliency, that is, is easily distinguished and contains few elements competing for attention. Simple diagram displays probably have higher saliency. This working memory difference in the ability to suppress bottom-up visual information may prevent low-spatial users from exerting more top-down effort on interpreting the problem features. In addition, the spatial properties in the diagram conditions may be consuming spatial working memory which is also used to build the mental model. In both cases, errant mental models are constructed more frequently in the diagram displays for low-spatial users and especially the simple diagram display.

Figure 8. Working memory model



LIMITATIONS OF THIS STUDY

A limitation of this study is that it did not include more robust measures of cognitive effort. Future studies might be benefit from using self-reports of cognitive effort such as the mental effort rating scale (Paas et al., 2003) to augment the technology adoption framework (Davis, 1989). For these types of problems, surveys of cognitive effort may be helpful for understanding user adoption behavior. In addition, in order to keep this study manageable, this study was not able to control other factors, such as testing the elimination of the "speed bump" in the simple diagram and text displays and testing out other visual treatments instead of diagrams like the iconic displays used in the Brase (2008) study.

The simple diagram and text conditions may have a ceiling effect. Since these displays, especially the diagram display for the high-spatial users, had a skewed distribution with 10 out of 10 correct as the most dominant score. Future studies that wish to test working memory limits may need to determine an optimal problem construction to avoid ceiling effects.

DESIGN CONSIDERATIONS

As the goal of this study was to help users with conditional probability problems in everyday, non-learning contexts, the designers of the user interface will need to take into account individual differences in spatial ability. Based on the findings, this study recommends the following guidelines:

1. Reduce relational complexity. With the exception of the simple diagram display for low-spatial users, simpler problems helped users produce more correct answers.

2. For low-spatial users, use simpler, text-only displays. These displays will facilitate performance better than diagrams will.

3. For high-spatial users, Venn and Euler-type diagrams can be safely used and can improve performance.

4. Use natural frequencies. All of the problems used in this study used natural frequencies and the best performing treatment, the simpler text-only displays, had a mean of 7.51 out of 10 correct. This is much higher than other studies performance measures.

5. If all components of the conditional probability problem must be displayed at once (e.g., the full contingency table representation), the diagrammatic display will provide the best performance overall.

6. If spatial ability measures are available for the target audience, designers can personalize the display based on spatial working memory capacity. Simpler diagram displays can be used with high-spatial users and simpler text-only displays can be used with low-spatial users.

Additional studies will need to be conducted in order to identify other design considerations, such as whether the intermediate step (with a question) in the simpler treatments used in this study) helps or hurts user performance while considering spatial ability; whether displays other than Venn and Euler-type diagrams might help (iconic displays, animation sequences, etc.); or whether low-spatial users need additional cognitive load when using diagrams.

IMPLICATIONS

The implications from this study are significant and broad. Manipulations of relational complexity and the use of Venn and Euler-type diagrams can aid in solving conditional probability problems. However, since low-spatial participants are impaired with the use of simple diagrams, and since this may be counterintuitive for many designers, it is likely that most use of simpler diagrams with less relational complexity may be adversely impacting close to half of their audiences. Analysis of the data from this study estimates that about 43% of the Venn diagram users (the low-spatial users who were negatively impacted in this study), could be at risk. While some research has been critical of the use of Venn diagrams for these problems (Brase, 2009) and some positive (Sloman, et al., 2003), none of the research has systematically tested Venn-type diagrams across a battery of problems while looking at working memory constraints.

Future research might fruitfully apply the theoretical models used for this study. With regard to conditional probability problems, relational complexity matters. Since these problems can be easily reduced into simpler forms with fewer potentially interacting elements, designers would be advised to consider simpler displays, especially when the audience for the displays cannot be easily dealt with individually. The extensive collection of research into people's performance in solving conditional probabilities and Bayesian reasoning has not yet applied concepts and frameworks from cognitive psychology like relational complexity and working memory. In addition, the recent research into the neural correlates of working memory and reasoning can provide linkages between the functional and physical descriptions of the mind. These frameworks can be applied to improve our understanding of user performance problems without having to refer to less testable constructs such as ‘frequency coding in the mind’ posited by frequentist interpretations of performance (Brase, 2009). Working memory, relational complexity, mental models and graph comprehension theories can be fruitful approaches for teasing out prescriptive details for improving problem performance with displays. While some prior research has looked at iconic displays for improving user performance in learning conditional probabilities, none have looked at the use of these displays from a graph comprehension, relational complexity and working memory perspective. For example, future studies could examine the relationship between relational complexity and the iconic displays previously found to help performance on conditional probability problems (Brase, 2009; Sloman, et al., 2003). Are iconic displays processed in working memory differently than diagrams? Such a study could identify additional visual display properties that make a cognitive difference.

As this study shows, individual differences in spatial ability do matter. While the most significant performance degradation with diagrams in this study was confined to low-spatial users, presentations of more complex conditional probability problems could potentially overwhelm high-spatial users as well. This would lead to a prediction that diagram displays could compete for spatial working memory resources and impair performance for even high spatial users, given a high level of complexity. This superior performance of text for low-spatial users in the face of high complexity might be more generalizable than this study. Other reasoning problems, such as syllogistic reasoning, predicate logic, mathematical reasoning or any problem-solving process in which numerous sets must be processed in parallel to determine relationships and appropriate inferences may also be impacted by complexity and diagrams. Applying relational complexity theory, theories of working memory, mental models theory and graph comprehension theory, as this study has, may be promising means for determining the effectiveness of computer displays on user performance for this and related problems.

Further improving real-world uses of diagrams

For automatically generated computer displays, like the ones used here, it would be useful to capture the users' spatial ability (or other cognitive factors) directly via a test of cognitive ability or infer it from other sources so that an appropriate display can be rendered. Based on graph comprehension research, it might be possible to correlate eye movements with inability to process diagrams effectively and hence, infer spatial ability. Also, some researchers have been developing simpler survey designs that can collect user preferences for visuals that correlate with measures of spatial ability (Blazhenkova & Kozhevnikov 2009). These techniques could be added to business intelligence and analytical software packages typically used in industry so that displays can be tailored appropriately based on individual differences in user spatial abilities.

For further improvement in solving these problems, other problem characteristics may need to be manipulated. While attributes of the problems like relational complexity and display type affected performance, other problem attributes may lurk that affect performance as well. In this study, problem texts varied in their word length. Since solving these problems involves reading and rereading text while processing sets, problems with longer word counts and larger number of potential sets to process may be more difficult to solve. A post-hoc repeated measures ANOVA test of the data in this study revealed that word count and number of sets affects performance (word count x the number of sets to process). Problems with higher word and set counts more 18% more errors than problems with lower word and set counts ($F(1,150)=10.66$, $p=.0014$). In addition, it is quite possible that lexical and syntactic complexity measures (see Lu, 2010) can help predict users' performance in these problems. While language and visual working memory capabilities are normally conceptualized as distinct and capable of independent processing, these working memory capabilities may be highly correlated with other measures of intelligence (McFarland, 2012). Increased syntactic complexity may increase working memory load (Fernandez-Duque, 2009). Therefore other tests of problem complexity and mental capabilities may also provide additional predictive capability. Lexical content may impede performance on some reasoning problems (Knauff & Johnson-Laird, 2002). Controlling lexical content and lexical and syntactic complexity may provide a fruitful path of inquiry for further improvement in user performance with these problems.

SUMMARY

This study has further clarified how computer-generated diagrams can (or cannot) aid in solving conditional probability problems. Across the suite of probability problems used in this study, users did better with diagrams than with text when the problem was more complex and did better with simple displays than complex ones when it was a text-only display. Diagrams facilitate solving complex conditional probability problems. However, spatial ability affects performance. Diagrams impede the reasoning ability for users with lower spatial ability. Prior research may have overlooked the interaction effect between spatial ability and reasoning. This interaction could be the source of some of the conflicting findings in the studies on the facilitation of conditional problem solving with diagrams. More importantly, in many everyday contexts where conditional probabilities are used and matter greatly, such as the conveying of health risk information and explanation of statistical evidence in legal proceedings, designers of displays should be cognizant of how diagrams help and hurt their audiences' ability to understand and solve these problems.

This study sought to identify design guidelines to help interface designers in constructing displays for conditional probability problems. The study found partial support for the study hypotheses concerning user performance. Complex diagram displays are superior in helping users generate correct answers than complex text displays and simple text displays are better than complex text displays. The study identified a significant interaction between users' spatial ability and the use of diagrams in which low-spatial users were impaired in their performance on simple diagrams compared with high-spatial users. The impact of this finding is broad. Designers of systems that convey conditional probability problems need to consider individual differences in spatial ability when using Venn and Euler-type diagrams.

REFERENCES

Andrews, G., Birney, D., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition, 24(6).* 1325-1340.

Awh, E., Barton B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science. 18(7),* 622-628.

Awh, E., Jonides, J., Smith, E.E., Schumacher, E.H., Koeppe, R.A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: evidence from PET. *Psychological Science. 7,* 25–31

Awh E., Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Science, 5(3),* 11-126.

Baddeley, A. (2012). Working Memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A. (2010). Working memory. *Current Biology, 20*, 4, 136-140.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews: Neuroscience. 4,* 829-839. doi:10.1038/nrn1201.

Baddeley, A. (2007). *Working Memory, Thought and Action.* Oxford University Press, New York, NY.

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Recent Advances in Learning and Motivation, 8*, 47–90. New York: Academic Press

Beck, D. M. & Kastner, S. (2009). To-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research, 49(10),* 1154-1165.

Blazhenkova, O. & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology, 23(5),* 638–663.

Birney, D. P., Halford, G. S., Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square task. *Education and Psychological Measurement. 66(1),* 146-171.

Boumans, M. (2008). Battle in the planning office: field experts versus normative statisticians. *Social Epistemology, 22(4),* 389-404.

Brase, G. L. (2008). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology,* DOI: 10.1002/acp.1460.

Brase G.L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*. DOI: 10.1002/bdm.421.

Brase, G. L. & Barbey, A. K. (2006). Mental representations of statistical information. *Advances in psychological research. (41).* 91-113.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: the role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127(1),* 3-21.

Brase, G. L., Fiddick, L. & Harries, C. (2006). Participant recruiting methods and statistical reasoning performance. *The Quarterly Journal of Experimental Psychology, 59(5),* 965-976.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In: *Studies in long-term memory,* Eds. A. Kennedy & A. Wilkes. Wiley.

Brooks, L. R. (1968). Spatial and verbal components in the act of recall. *Canadian Journal of Psychology. (22).* 349–368

Calvillo, D. P., DeLeeuw, K., Revlin, R. (2006). Deduction with Euler Circles, Diagrams that hurt. *Diagrams 2006, LNAI 4045,* 199-203.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4(2),* 75-100.

Chase, W & Simon, H. A. (1973). The mind eye's in chess. In: *Visual information processing,* ed. by W. G. Chase. Academic Press.

Cleeremans, A. & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–53.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin,112(1),* 155-199.

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement, 7(3),* 249-253.

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*. 58, 1-73.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W.S. Sossin, J.-C. Lacaille, V.F. Castellucci & S. Belleville (Eds.). *Progress in Brain Research, Vol. 169.* Elsevier B.V.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-185.

Cowan, N., Nugent, L. D., Elliot, E. M., Ponomarev, I. & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70,* 1082–97.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42-100.

Crosby, M. A., Iding, M. K., Chin, D. N. (2003). Research on task complexity as a foundation for augmented cognition. *Proceedings of the 36th Hawaii International Conference on System Sciences, 2003*.

Darwin, C. J., Turvey, M. T. & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3,* 255–67.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, September.*

den Heyer, K.; Barrett, B. (1971). Selective loss of visual and verbal information in STM by means of visual and verbal interpolated tasks. *Psychonomic Science*, *25*, 100-102

Dennis, S., McArthur, R., & Bruza, P. (1998). Searching the world wide web made easy? The cognitive load imposed by query refinement mechanisms. *Proceedings of the 3rd Australian Document Computing Symposium (ADCS '98)*, 65–71.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic & A. Tversky (Eds), *Judgment under uncertainty: Heuristics and biases.* (pp. 249-267). Cambridge, England: Cambridge University Press.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). Manual for kit of factor-referenced cognitive tests. Princeton: Educational Testing Service.

Ericsson, K. A. (1985). Memory skill. *Canadian Journal of Psychology, 39,* 188–231.

Ericsson, K. A., Chase, W. G. & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208,* 1181–82.

Evans, J., Handley, S. J., Perham, N. Over, D.E., Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*. *77*, 197-213.

Fangmeier, T., Knauff, M., Ruff, C., & Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience, 18(3),* 320-334.

Feldman Barret, L., Tugade, M. M., Engle, R. W. (2004). Individual differences in working memory capacity and dual-processing theories of the mind. *Psychological Bulletin, 130(4)*, 555-573.

Fernandez-Duque, D. (2009). Cognitive and neural underpinnings of syntactic complexity. In T. Givon & M. Shibatani (Eds), *The Genesis of Syntactic Complexity*, 433-460. J. Benjamin Publisher.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology, 16,* 449–84.

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & H. Narayanan (Eds). *Diagramatic representation and inference*, 18-31. Berlin: Springer-Verlag.

Glanzer, M. & Razel, M. (1974). The size of the unit in short-term storage. *Journal of Verbal Learning and Verbal Behavior, 13,* 114–31.

Gigerenzer, G. & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *BMJ*. 327. 741-744.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review.* 102. 684-701.

Gigerenzer G. & Hoffrage, U. (1999). Overcoming difficulties in Bayesian Reasoning: A reply to Lewis & Keren and Mellers & McGraw. *Psychological Review*. 106. 425-430.

Girotto, V. & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition*. 78. 247-276.

Graesser II, A. & Mandler, G. (1978). Limited processing capacity constrains the storage of unrelated sets of words and retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 86–100.

Hale, S., Myerson, J., Emery, L.J., Lawrence, B. M., Dufault, C. (2008). Variation in working memory across the life span. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 194–224). New York: Oxford University Press.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science, 16(1),* 70-76.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Science, 11(6)*, 236-242.

Halford, G.S., Wilson, W.H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, (21),* 803-865.

Halford, G.S., Wilson, W.H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14(11),* 497-505.

Halford, G. S., Mayberry, M. T. & Bain, J. D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory and Cognition,* 16, 480–87.

Hegarty, M. (1997). Individual differences in use of diagrams as memory in mechanical reasoning. *Learning and Individual Differences, 9(1),* 19-42.

Hegarty, M. (2000). Capacity limits in diagrammatic reasoning. *Diagrams 2000, LNAI 1889,* 194-206.

Hegarty, M. & Waller, D.A. (2005) Individual differences in spatial ability. In P.. Shah & A. Miyake (Eds), *The Cambridge Handbook of Visuospatial Thinking.* (pp. 121-169). Cambridge, England: Cambridge University Press.

Houdé, O., & Tzourio-Mazoyer, N. (2003) Neural foundations of logical and mathematical cognition. *Nature Reviews, Neuroscience 4*, 507-514

Hoffman, M. H. G. (2007). Cognitive conditions of diagrammatic reasoning. Working paper. Accessed February 13, 2010. https://smartech.gatech.edu/bitstream/1853/23809/1/wp24.pdf

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, 84, 343–352.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V. & Legrenzi, M.S. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*. 106. 62-88.

Jones, D., Farrand, P., Stuart, G. & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1008–18.

Jonides, J. Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G. & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193-224

Jonides J., Smith E. E., Koeppe R. A., Awh E., Minoshima S., Mintun M.A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363*, 623–25.

Joseph, S, Yukawa, J., Suthers, D., & Harada, V. (2006). Searching emergent vocabularies: exploring methods to reduce cognitive load during web navigation and resource contribution. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06),Vol. 7*.

Kahnamen, D. & Lovallo, D. (1993). Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science, 39(1),* 17-31.

Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–94.

Klauer, K.C., Stegmaier, R. & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking and Reasoning, 3(1).* 9-47.

Klauer, K. C., & Zhao, Z. (2004). Double disassociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133(3),* 355-381.

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition and Computation, 9,* 109-137.

Knauff, M. (2007). How our brains reason logically. *Topoi, 26(1),* 19-36. doi: 10.1007/s11245-006-9002-8

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition & Computation, 9,* 109-137. doi:10.1080/13875860902887605.

Knauff, M., Jola, C. & Strube. (2001). Spatial reasoning: no need for visual information. In D.R. Montello (Ed.), *COSIT 2001, LNCS 2205*, 447−457.

Knauff, M. & May, E. (2006). Visual Imagery in Deductive Reasoning: Results from experiments with sighted, blindfolded, and congenitally totally blind persons. *The Quarterly Journal of Experimental Psychology, 59(1),* 161-177.

Knauff, M. & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, *30,* 363-371.

Kroger, J.K., Sabb, F. W., Fales, C. I., Bookheimer, S. Y., Cohen, M. S., Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebral Cortex, 12*, 477-485.

Kulpa, Z, (1994). Diagrammatic representation and reasoning. *Machine Graphics & Vision. 3(1/2),* 77-103

Kurzenhäuser, S., & Hertwig, R. (2006). How to foster citizens' statistical reasoning: Implications for genetic counseling. *Community Genetics, 9,* 197-203.

Kurzenhäuser, S. & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher*. 24. 516-521.

Larkin, J. H., & Simon, H.A. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science, 11(1),* 65-99.

Lo, S. & Helander, M. G. (2004). Developing a formal usability analysis method for consumer products. *Proceedings of ICAD2004, The Third International Conference on Axiomatic Design Seoul – June 21-24*.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review 95,* 492–527.

Logan, G. D. & Klapp, S. T. (1991). Automatizing alphabet arithmetic: Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–95.

Logie, R. H. (2011). The functional organization and capacity of working memory. *Current Directions in Psychological Science, 20(4),* 240-245.

Lu, X. (2010).Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15(4),* 474-496.

Macchi, L. (2000). Partitive formulation in probabilistic problems: beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes. 82(2),* 217-236.

Mandler, G. & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22.

Masri, K., Parker, D., Gemino, A. (2008). Using iconic graphs in entity-relationship diagrams: the impact on understanding. *Journal of Database Management, 19(3).*

Dennis J. McFarland (2012). A single g factor is not necessary to simulate positive correlations between cognitive tests. *Journal of Clinical and Experimental Neuropsychology, 34(4)*, 378-384.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist,32(1)*,1-19.

McCains, S. & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *The Journal of Neuroscience*. *312(2)*, 587-597.

McKone, E. (1995). Short-term implicit memory for words and nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1108–26.

Mellers, B. A. & McGraw, A. P. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological review. 106.* 417-424.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Miller, G.A., Galanter, E. and Pribram, K.H. (1960). *Plans and the structure of behavior.* Holt, Rinehart and Winston, Inc., New York

Miyake, A. & Shah, P. (1999). An Introduction. In A. Miyake & P. Shah (Eds.) *Models of Working Memory.* Cambridge University Press.

Miyake, A., Friedman, N. P., Rettinger, D.A., Shah, P., Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130(4),* 621-640.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory and Cognition, 19,* 332–40.

Nelson, W., Reyna, V., Fagerlin, A., Lipkus, I. & Peters, E. (2008). Clinical implications of numeracy: theory and practice. *American Behavioral Medicine*. *35.* 261-274.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in Working Memory*. New York: Oxford University Press.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429-434.

Paas,F., Tuovinen, J., Tabbers, H. Van Gerven, P. (2003): Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38(1),* 63-71

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 45*, 255–287

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Paulos, J.A. (1988). *Innumeracy: Mathematical Illiteracy and Its Consequences*. New York: Vintage Books.

Pollack, I., Johnson, I. B. & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*, 137–46.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12,* 61-86.

Pylyshyn, Z., Burkell, J., Fisher, B., Sears, C., Schmidt, W. & Trick, L. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology, 48*, 260–83.

Raaijmakers, J. G. W. & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed capacity models of visual working memory. *Proceedings of the National Academy of Sciences, 105(16),* 5975-5979.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common and distinct neuronal processes. *Neuropsychologia, 41,* 1241-1253.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of groups. *Quarterly Journal of Experimental Psychology, 21*, 137–47.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in working memory. *Memory & Cognition, 34(2), 344-355.*

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General. 136(4),* 663-684.

Schmutz, P., Heinz, S., Métrailler, Y., & Opwis, K. (2009). Cognitive load in eCommerce applications—measurement and effects on user satisfaction. *Advances in Human-Computer Interaction, 2009.*

Schnotz, W. & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19,* 469-508.

Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task wisely and learn by doing. *Instructional Science*. 28. 227-262.

Shah, P. (2002). Graph comprehension: The role of format, content and individual differences. In M. Anderson, B. Meyer, & P. Oliver (Eds) *Diagrammatic Reasoning and Representation.* Springer-Verlag.

Simon, H. A. (1974). How big is a chunk? *Science, 183,* 482–88.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes.* 91. 296-301.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* Whole No. 498.

Stenning, K. & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science 19(1),* 97-140.

Suo, X, Zhu, Y., Owen, S. (2008). A task centered framework for computer security data visualization. In J.R. Goodall, G. Conti, and K.-L. Ma (Eds.): *VizSec 2008, LNCS 5210,* 87–94.

Sweller, J. (1998). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12,* 257-285.

Toplak, M.E., West, R.F., & Stanovich, K.E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristic-and-biases tasks. *Memory Cognition 39*, 1275-1289

Trbovivh, P. L., & LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition 31(5)*, 738-745.

Trick, L. M. & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 331–51.

Ungerleider L.G., Haxby J. V. (1994). “What” and “where” in the human brain. *Current Opinion in Neurobiology. 4,* 157–65

Van der Stegchel, S, Belopolsky, A. V., Peters, J.C. Wijnen, J. G, Meeter, M., Theeuwes, J. (2009). The limits of top-down control of visual attention. *Acta Psychologica*, *132,* 201-212.

Vandierendonck, A., & De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *Quarterly Journal of Experimental Psychology, 50A(4)*, 803–820.

Van Merriënboer, J. & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education, 44,* 85-93.

Van Opstal, F., Gevers, W., De Moor, W. & Verguts, T. (2008). Dissecting the symbolic distance effect: comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin and Review.* doi: 10.3758/PBR.15.2.419.

Venkatesh, V., Morris, M. G. & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly, 27(3).* 425-478.

Wager. T.D. & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 255-274

Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin, 81,* 695–711.

Waugh, N. C. & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104.

Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of Experimental Psychology, 68*, 413–19.

Xing, J. (2004). Measures of information complexity and the implications for automation design. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-04/17.

Xing, J. (2007). Information complexity in air traffic control displays. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-07/26

Xing, J. (2008). Designing questionnaires for controlling and managing information complexity in visual displays. Washington, DC: FAA Office of Aerospace Medicine, Report No. DOT/FAA/AM-08/18.

Zago, L., & Tzourio-Mazoyer. (2002). Distinguishing visuospatial working memory and complex mental calculation areas within the parietal lobes. *Neuroscience Letters (331)*, 45-49.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science. 21(2).* DOI: 10.1207/s15516709cog2102_3.

Zhang, J. & Norma, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science. 18(1).* 87-122.

Zhang, G. & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition, 13,* 193–201.

Zhu, L. & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representations in mental computations. *Cognition, 98,* 298-308.

Zimmer, H.D., Münzer, S., & Umla-Runge, K. (2010). Visuo-spatial working memory as a limited resource of cognitive processing. In M.W. Crocker, and J. Siekmann (eds.), *Resource-Adaptive Cognitive Processes*, Springer-Verlag Beling Heidelberg.

APPENDIX A. BACKGROUND QUESTIONS

To be asked first at the beginning of the study.

Age: _____

Gender: ○ Male ○ Female

What is your highest educational background?
○ High School ○ Undergraduate ○ Graduate School ○ PhD

If you have a degree in progress, what is it?
○ None ○ High School ○ Undergraduate ○ Graduate School ○ PhD

Concentration:________________________________

Is English your native language? ○ Yes ○ No

Are you visually impaired? ○ Yes ○ No

Have you had experience with probability theory or statistics? ○ Yes ○ No

How would you describe your skill in probability theory or statistics?

① ② ③ ④ ⑤ ⑥ ⑦
No skill Expert
skills

APPENDIX B. SPATIAL ABILITY TEST

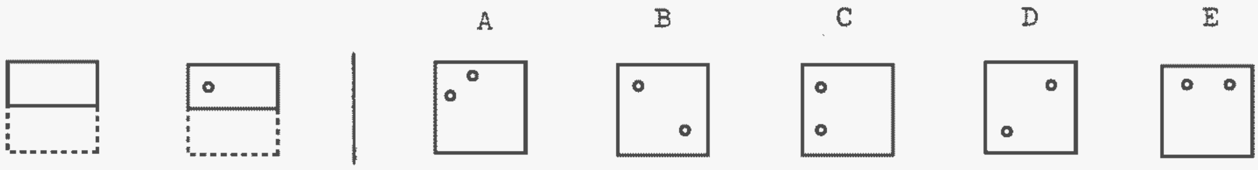
(This page is intentionally left blank)

Name ____

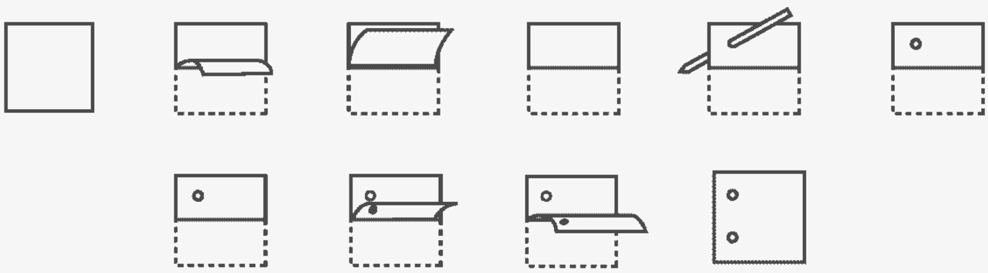
PAPER FOLDING TEST — VZ-2

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



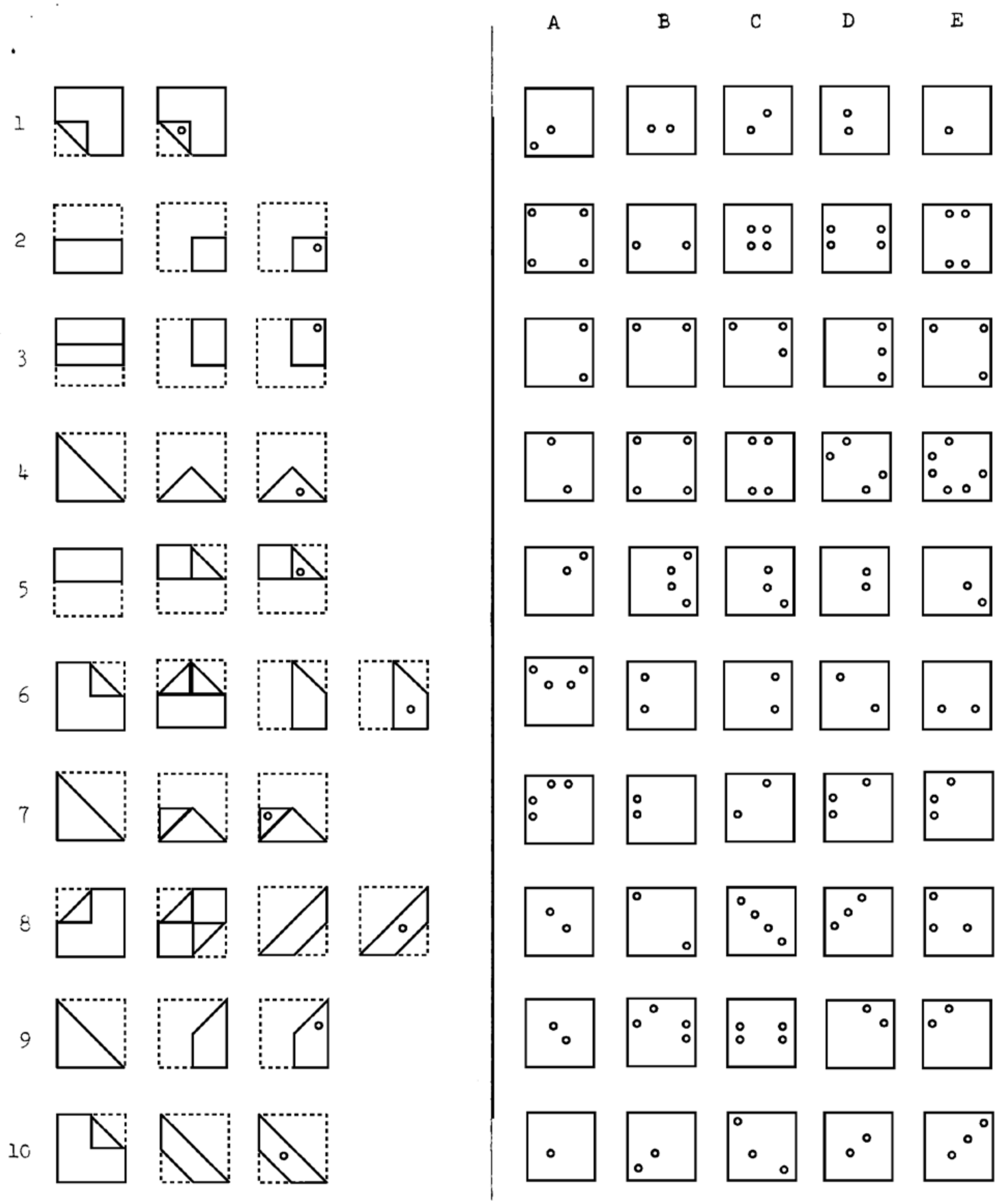
In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

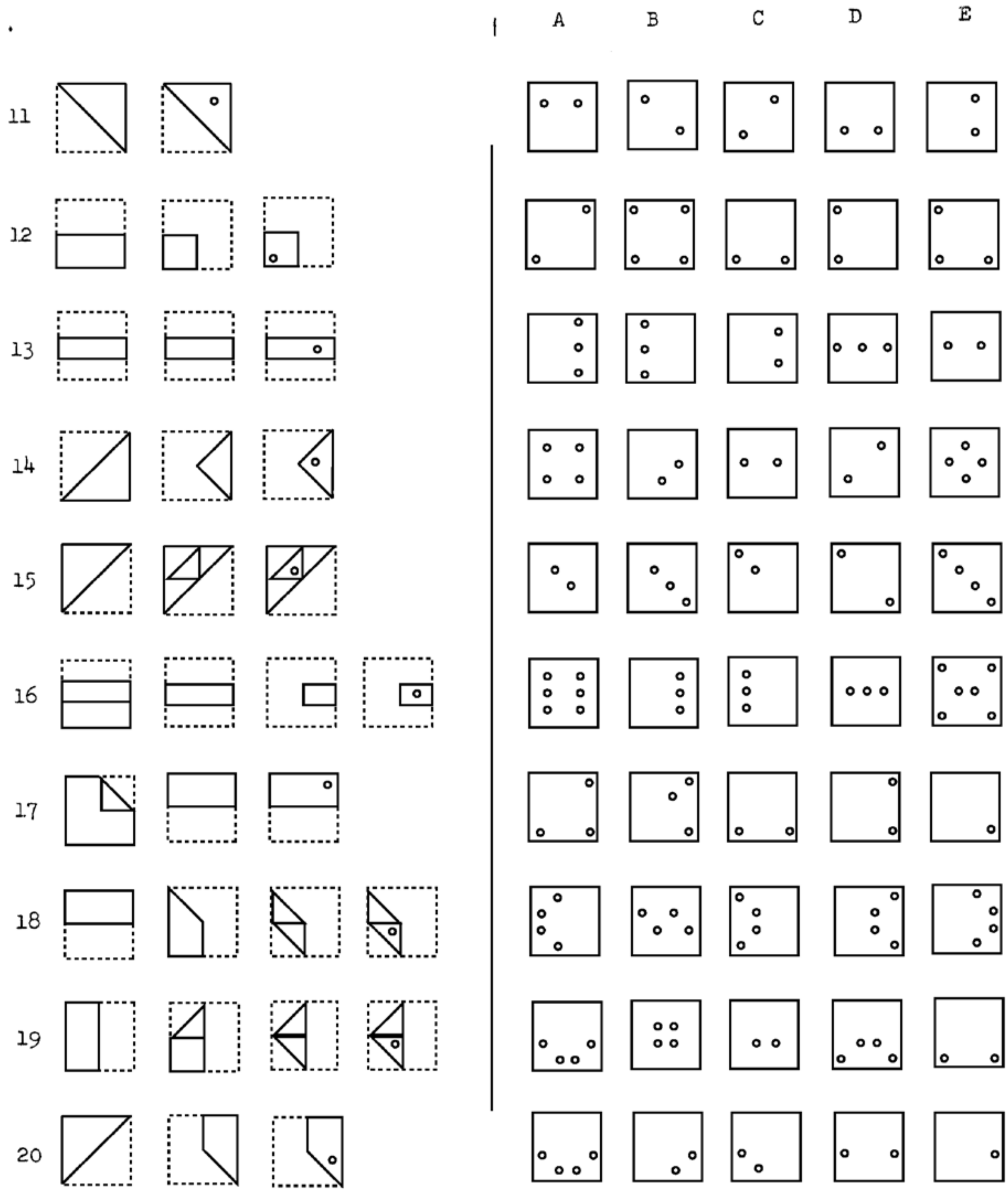
Part 1 (3 minutes)



DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Part 2 (3 minutes)



DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. **STOP.**

APPENDIX C. SATISFACTION INSTRUMENT

<u>After completion of all problems</u>

1-I am confident that all my answers are correct.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

2-I am unsure that the approach I took in solving these problems is correct.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

3-I am confident that I will be able to solve problems like this in the future.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

4-The presentation of these problems was useful for comprehension tasks.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

5-Learning to use this application was easy.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

6-The presentation of the problems enhanced the effectiveness of my comprehension.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

7-I have the knowledge and ability to use this application.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

8-The presentation of the problems made comprehension tasks easier.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

9-The presentation of the problems was clear and understandable.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

10-Remembering how to perform tasks was easy.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

11-I felt in control when using this application.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

12-The presentation of the problems helped me find answers quickly.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

13-It was easy to use this application.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

14-The comprehension task on this presentation required less mental effort.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

15-I felt frustrated when performing the tasks for these problems.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

16-This presentation of the problems improved my understanding of the problems.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

17-In general, I am satisfied with the presentation of the problems.

①	②	③	④	⑤	⑥	⑦
Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree

APPENDIX D. DATA ANALYSIS ASSUMPTIONS

Dependent variable analysis discovered some non-normal data with non-equal variances (heteroskedasticity) that could not be adequately resolved with transformations. However, sample sizes in each cell were large (>35) and all dependent variables had well bounded variances due to the responses being limited to the Likert scales for the satisfaction measure and the number correct (between 0 and 10). The distribution for correct, ease of use and satisfaction are shown in Figures 9, 10 and 11). The residuals for correct were non-normal and correct had non-equal variances. The residuals for perceived ease of use and perceived usefulness were non-normally distributed, but both variables had equal variances.

Figure 9. Distribution of Correct.

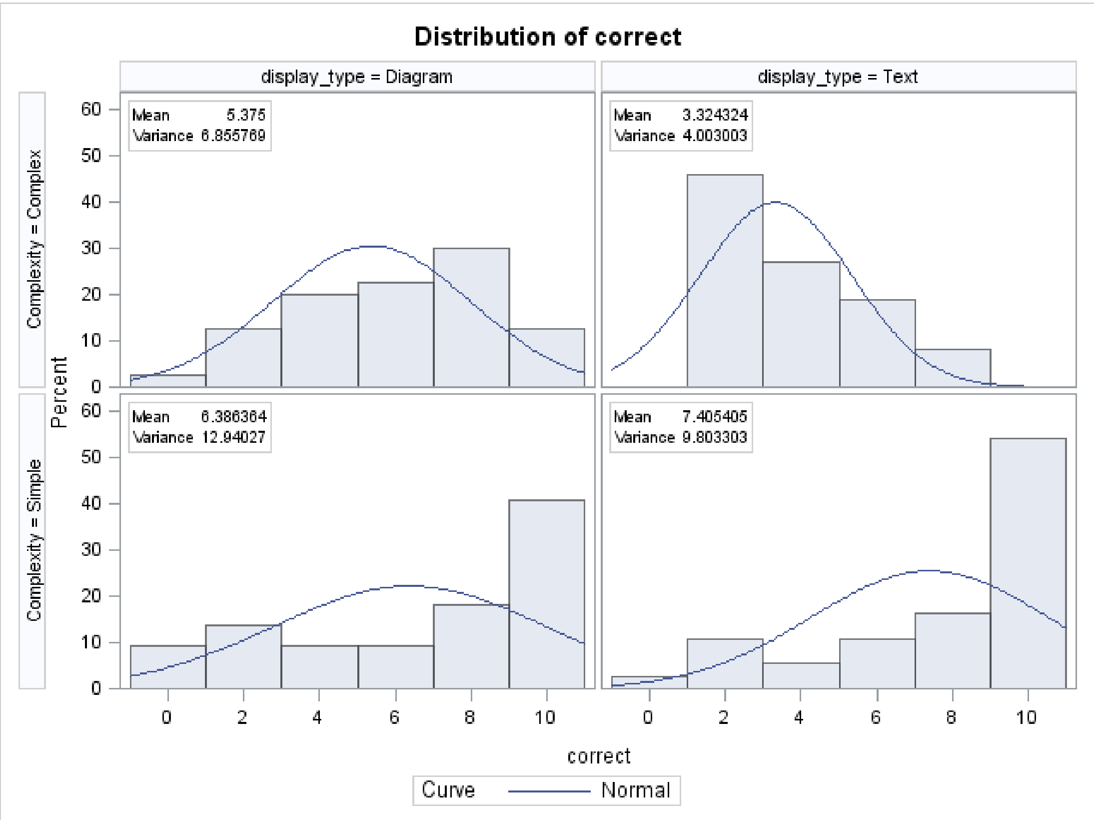


Figure 10. Distribution of satisfaction - perceived ease of use

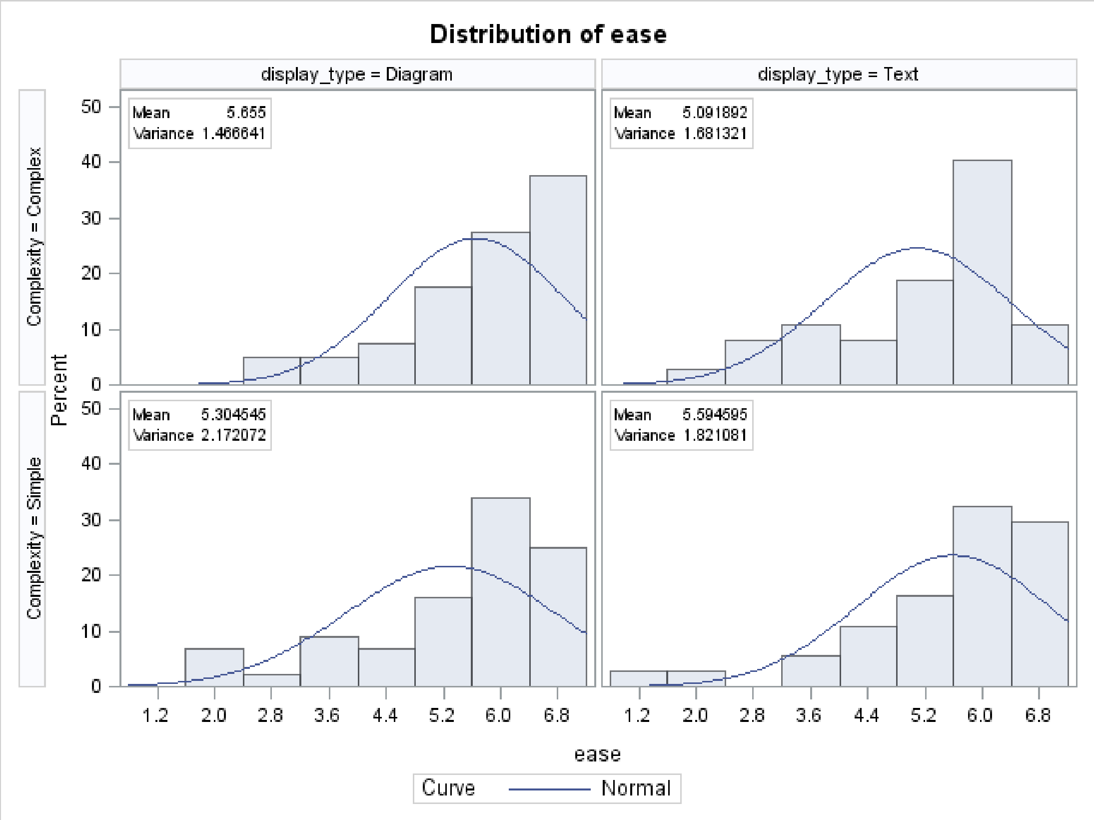
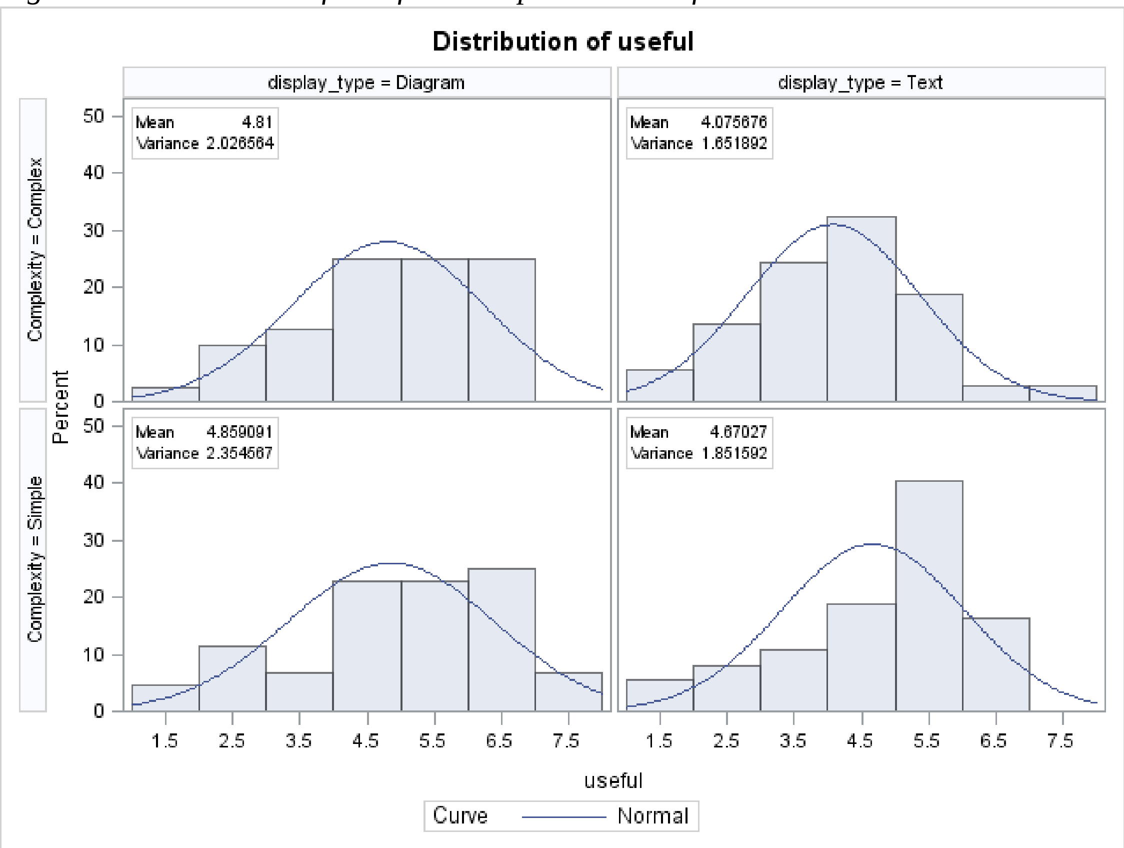


Figure 11. Distribution of satisfaction – perceived usefulness



ANCOVA ASSUMPTIONS

Since the covariate, spatial ability, was measured before the beginning of the experiment, the covariate was measured independent of the treatments. Further analysis of the variable correct, ease and perceived usefulness revealed that the distributions of residuals are not normal and for correct, that the variances in all of the four conditions are not equal. These heteroskedasticity and non-normality issues with correct could not be addressed with conventional transformation and all analysis was conducted without transformation of dependent. Table 15 shows the results of relevant the test statistic for each measure.

Table 15. Dependent measures of normality and heteroskedasticity

Variable	Shapiro-Wilk test for normality	Levene test for heteroskedasticity
correct	W=.9742, p=.0047	F(3)=5.84, p=.0008
perceived ease of use	W=.9199, p<.0001	F(3)=0.46, p=.7136
perceived usefulness	W=.9638, p=.0004	F(3)=0.59, p=.6235

T-TEST ASSUMPTIONS FOR PERFORMANCE AND SATISFACTION

For t-tests, wherever non-equal variances were encountered with correct, the Satterthwaite approximate t test was used to report the *t* and the p values. Since sample sizes were large, normality was dealt with as in the ANCOVA analysis and variables were not transformed. The folded F statistic was used to compare variances in each of the t-tests. Table 16 lists the folded F statistic p value for each comparison used.

Table 16. Folded F Test for equality of variances p-values

Test	Correct	Ease	Useful	Difficulty
6D>6T	0.1062	0.6747	0.5382	0.3213
3D>3T	0.3958	0.5913	0.4627	0.9590
3D>6D	0.0469	0.2168	0.6377	0.7400
3T>6T	0.0086	0.8120	0.7339	0.5379

One comparison with non-equal variances had a significant difference in means (3T>6T for correct). The significance of this comparison of these two conditions' means was p<.0001.

APPENDIX E. VALIDATION OF THE SATISFACTION INSTRUMENT

The satisfaction survey (see Appendix A) was analyzed to determine its reliability.

1. Factor Analysis

(Questions 6, 11, 12 and 14 were dropped because of their low loadings.)

	Factor 1 (Perceived Usefulness)	Factor 2 (Perceived Ease of Use)	Item Communality
Question 1	0.72		0.67
Question 2		0.84	0.77
Question 3	0.87		0.83
Question 4		0.82	0.75
Question 5	0.82		0.80
Question 7		0.55	0.71
Question 8		0.80	0.80
Question 9	0.76		0.83
Question 10		0.77	0.70
Question 13	0.80		0.71

2. Internal Consistency of the Instrument

Variable	Number of Items	Cronbach's α Value
Perceived usefulness	5: Q1,Q3,Q5,Q9,Q13	0.92
Perceived ease of use	5: Q2,Q4,Q7,Q8,Q10	0.90

APPENDIX F. BACKGROUND INFLUENCE ON PERFORMANCE

Table 17 shows the analysis of participant background on performance. None of the background survey questions significantly influenced performance.

Table 17. Participant background influence on performance

Variable	*F* test	*p* value
Age	0.00	.9831
English	0.32	.5734
Gender	1.15	.2851
Probability skill	1.74	.1897
Probability experience	1.03	.3128
Degree attained	1.08	.3002
Degree in progress	0.58	.4468
Visual impairment	2.25	.1357

APPENDIX G. TREATMENTS

PROBLEM 1. BLUE CAB

Prevalence: 15% (15/100)
Sensitivity: 80% (12/15)
Specificity: 80% (68/85)

H = Cab is Blue Not H = cab is Green
D = Cab is seen as Blue Not D = cab is seen as Green

	H	Not H	
D	12	17	29
Not D	3	68	71
	15	85	100

Answer final: 12 / 29 = 41.379% * 20 = 8.2758 (8 or 9)
Answer intermediate: 29 / 100 = 29% * 50 = 14.5, (14 or 15)

<u>Text RC6</u>

T1: A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. When the cabs were really Blue, the witness said they were Blue in 12 out of 15 tests. When the cabs were really Green, the witness said they were Blue in 17 out of 85 tests. Overall, the witness identified 29 cabs as Blue.
Question: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

<u>Diagram RC6</u>

T1: A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue.
T2: The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. The results are as follows:
Right circle label: Green cabs (85)
Left circle label: Blue cabs (15)
Ellipse: Cabs witness said were Blue (29)
Ellipse right: Green cabs witness said were blue (17)
Ellipse left: Blue cabs witness said were blue (12)
Question: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

<u>Text RC3</u>

T1: A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. In 29 of those tests, the witness said the cab was Blue.
Question A: If 50 accidents occurred over the next year, how many times would this witness say the cab was Blue?
T2: In the tests in which the witness said the cab was Blue, the cab was actually Blue 12 times.
Question B: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

Diagram RC3

T1: A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate cabs in the city. On the night of the accident, a witness said the cab was Blue.
T2: The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs by conducting 100 tests on the witness. The results are as follows:
Circle A: Cabs tested (100)
Circle B: Cabs witness said were Blue (29)
Question A: If 50 accidents occurred over the next year, how many times would this witness say the cab was Blue?
Circle C: Cabs actually Blue (12)
Question B: If 20 accidents occurred over the next few months in which the witness said the cab was Blue, how many accidents would a Blue cab actually be involved in?

PROBLEM 2. MAMMOGRAPHY

Prevalence: .8% (8/1000)
Sensitivity: 80% (8/10)
Specificity: 90.4% (895/990)

H = Has disease — Not H = does not have disease
D = Tested positive — Not D = Tested negative

	H	Not H	
D	8	95	103
Not D	2	895	897
	10	990	1000

Answer final: 8 / 103 = 7.76% * 80 = 6.21 (6 or 7)
Answer intermediate: 103 / 1000 = 10.3% * 750 = 77.25 or 103 / 1000 = 10% * 750 (77 or 78, 75)

Text RC6:

T1: A doctor runs a clinic specializing in breast cancer. Over the last four years, the doctor has seen 1,000 women for routine screening. 103 out of those 1,000 women had tests results showing cancer and 897 with test results indicating no cancer. Overall, 10 women actually had the disease with 8 women who had test results showing cancer and 2 with test results showing no cancer.
Question: Based on this data, if another 80 women had positive tests, how many would you expect to have the disease?

<u>Diagram RC6:</u>

T1: A doctor runs a clinic specializing in breast cancer.
T2: Over the last four years, the doctor has seen many women for routine screening. The results are as follows:
Right circle label: Women with tests results showing no cancer (897)
Left circle label: Women with tests results indicating cancer (103)
Ellipse: Women with the disease (10)
Ellipse right label: Women with positive tests who have the disease (8)
Ellipse left label: Women with negative tests who have the disease (2)
Question: Based on this data, if another 80 women had positive tests, how many can be expected to have the disease?

<u>Text RC3</u>

T1: A doctor runs a clinic specializing in breast cancer. Over the last four years, the doctor has seen 1,000 women for routine screening. 103 out of those 1,000 women had tests results showing cancer.
Question A: Based on the data from this initial screening of women, if another 750 women were screened, how many can be expected to have positive results?
T2: Of those who tested positive, 8 women actually had the disease.
Question B: Based on this data, if another 80 women had positive tests, how many can be expected to have the disease?

<u>Diagram RC3:</u>

T1: A doctor runs a clinic specializing in breast cancer.
T2: Over the last four years, the doctor has seen many women for routine screening. The results are as follows:
Circle A label: Women screened (1000)
Circle B label: Women with positive tests (103)
Question A: Based on the data from this initial screening of women, if another 750 women were screened, how many can be expected to have positive results?
Circle C: Women with the disease (8)
Question B: Based on this data, if another 80 women had positive tests, how many can be expected to have the disease?

PROBLEM 3. STUDENTS WITH GLASSES

Prevalence: 90% (90/100)
Sensitivity: 50% (45/90)

Specificity: 70% (7/10)

H = Person is a student   Not H = Person is not a student
D = Person wears glasses   Not D = Person does not wear glasses

	H	Not H	
D	45	3	48
Not D	45	7	52
	90	10	100

Answer final: 45 / 48 = 93.75% * 20 = 18.75 (18 or 19)
Answer A: 48 / 100 = 48% * 50 = 24.0

<u>Text RC6:</u>

T1: On a university campus are people who are students and non-students, some who wear glasses and some who do not. 90 out of every 100 people you meet are students of this university. Of the 90 students, 45 wear glasses. Of the remaining 10 people that are not students, 3 also wear glasses. Overall, 48 people on campus wear glasses.
Question: Suppose you meet a group of 20 people who wear glasses on the campus. How many of them would you expect to be students at this university?

<u>Diagram RC6:</u>

T1: On a university campus are people who are students and non-students, some who wear glasses and some who do not.
T2: A recent count of a sample population is as follows:
Left circle label: University students (90)
Right circle label: Not university students (10)
Ellipse: People wearing glasses (48)
Ellipse left: Students wearing glasses (45)
Ellipse right: Non-students wearing glasses (3)
Question: Suppose you meet a group of 20 people who wear glasses on the campus. How many of them would you expect to be students at this university?

<u>Text RC3:</u>

T1: On a university campus are people who are students and non-students, some who wear glasses and some who do not. 48 out of every 100 people you meet wear glasses.
Question A: Suppose you were to meet 50 people on campus. How many of them would you expect to be wearing glasses?
T2: Of those people on campus wearing glasses, 45 of them are students.
Question B: Suppose you were to meet a group of 20 people who wear glasses on campus. How many of them would you expect to be students at this university?

<u>Diagram RC3:</u>

T1: On a university campus are people who are students and non-students, some who wear glasses and some who do not.
T2: A recent count of a sample population is as follows:

Circle A label: People on campus (100)
Circle B label: People wearing glasses (48)
Question A: Suppose you were to meet 50 people on campus. How many of them would you expect to be wearing glasses?
Circle C label: Students wearing glasses (45)
Question B: Suppose you were to meet a group of 20 people who wear glasses on campus. How many of them would you expect to be students at this university?

PROBLEM 4. WORKFORCE PLANNING

Prevalence: 70% (28/40)
Sensitivity: 14% (4/28)
Specificity: 33% (4/12)

H = Prefers downtown location
Not H = Prefers suburban location
D = Prefers night shifts
Not D = Prefers day shifts

	H	Not H	
D	4	8	12
Not D	24	4	28
	28	12	40

Answer final: 4 / 12 = 33.33% * 75 = 25 (24)
Answer A: 12 / 40 = 30% * 200 = 60

<u>Text RC6:</u>

T1: A manufacturing company asked every person who was hired if they prefer to work during the day or at night and if they prefer to work at the company's downtown location in the city's center or if they prefer to work at the company's suburban location. The company found that out of 40 employees surveyed, 28 prefer working days and 12 preferred working nights. A total of 29 preferred working downtown. Of those 29, 4 expressed an interest in working nights at the downtown location and 24 expressed an interest in working days at the downtown location.

Question: If the company hires another 75 employees who prefer working nights, how many would you expect to prefer the downtown location?

<u>Diagram RC6:</u>

T1: A manufacturing company asked every person who was hired if they prefer to work during the day or at night and if they prefer to work at the company's downtown location in the city's center or if they preferred to work in the company's suburban location.
T2: The results are as follows:
Left circle label: Prefers nights (12)
Right circle label: Prefers days (28)
Ellipse: Prefers downtown (29)
Ellipse left: Prefers downtown nights (4)

Ellipse right: Prefers downtown days (24)
Question: If the company hires another 75 employees who prefer working nights, how many would you expect to prefer the downtown location?

<u>Text RC3:</u>

T1: A manufacturing company asked every person who was hired if they prefer to work during the day or at night and if they prefer to work at the company's downtown location in the city's center or if they preferred to work in the company's suburban location. The company found that out of 40 employees surveyed, 12 preferred working nights.
Question A: If the company hires another 200 employees, how many would you expect to prefer working nights?
T2: Of those surveyed who said they preferred working nights, 4 said they preferred working nights downtown.
Question B: If the company hires another 75 employees who prefer working nights, how many would you expect to prefer the downtown location?

<u>Diagram RC3:</u>

T1: A manufacturing company asked every person who was hired if they prefer to work during the day or at night and if they prefer to work at the company's downtown location in the city's center or if they preferred to work in the company's suburban location.
T2: The results are as follows:
Circle A label: People asked (40)
Circle B label: Prefers nights (12)
Question A: If the company hires another 200 employees, how many would you expect to prefer nights?
Circle C label: Prefers nights downtown (4)
Question B: If the company hires another 75 employees who prefer working nights, how many would you expect to prefer the downtown location?

PROBLEM 5. COOKIES

Prevalence: 20% (10/50)
Sensitivity: 70% (7/10)
Specificity: 70% (28/40)

H = Cookie is salty Not H = cookie is sweet
D = Cookie is round Not D = Cookie is square

	H	Not H	
D	7	12	19
Not D	3	28	31
	10	40	50

Answer final: 7 / 19 = 36.84% * 10 = 3.684 (3 or 4)
Answer A: 19 / 50 = 38% * 20 = 7.6 (7 or 8)

Text RC6:

T1: There is a bag of 50 sweet or salty cookies with various kinds of shapes. In the bag, 10 of the cookies are salty and 40 are sweet. A total of 19 cookies are round with 7 of the round being salty cookies, and 12 of the round being sweet cookies.
Question: Imagine you taking 10 round cookies from the bag of 50. How many of them would you expect to be salty?

Diagram RC6:

T1: There is a bag of sweet or salty cookies with various kinds of shapes.
T2: The contents of the bag are as follows:
Right circle label: Salty (10)
Left circle label: Sweet (40)
Ellipse: Round (19)
Ellipse right: Salty round (7)
Ellipse left: Sweet round (12)
Question: Imagine taking 10 round cookies from the bag of fifty. How many of them would you expect to be salty?

Text RC3:

T1: There is a bag of 50 sweet or salty cookies with various kinds of shapes. Of the different cookies in the bag, 19 are round.
Question A: Imagine taking out 20 cookies from the bag. How many would you expect to be round?
T2: Of the round cookies in the bag, 7 cookies are salty.
Question B: Imaging taking 10 round cookies from the bag of fifty. How many would you expect to be salty?

Diagram RC3:

T1: There is a bag of sweet or salty cookies with various kinds of shapes.
T2: The contents of the bag are as follows:
Circle A label: Cookies (50)
Circle B label: Round cookies (19)
Question A: Imagine taking out 20 cookies from the bag. How many would you expect to be round?
Circle C label: Salty round cookies (7)
Question B: Imaging taking 10 round cookies from the bag of fifty. How many would you expect to be salty?

PROBLEM 6. COLOR-BLIND MEN

Prevalence: 50% (500/1000)
Sensitivity: 5% (25/500)
Specificity: 99.6% (498/500)

H = Person is male	Not H = Person is female
D = Person is color-blind	Not D = Person is not color-blind

	H	Not H	
D	25	2	27
Not D	475	498	973
	500	500	1000

Answer final: 25 / 27 = 92.592% * 10 = 9.259 (9 or 10)
Answer A: 27 / 1000 = 2.7% * 100 = 2.7 (2 or 3)

<u>Text RC6:</u>

T1: A small town has a population of 1,000 people (500 men and 500 women). A total of 27 people in the town are color-blind with 2 women who are color-blind and 25 men who are color-blind.
Question: If you randomly selected 10 people from all the color blind people in this town, how many would you expect to be men?

<u>Diagram RC6:</u>

T1: A small town has a population of both women and men, some of whom are color-blind.
T2: The population can be depicted as follows:
Right circle label: Men (500)
Left circle label: Women (500)
Ellipse: Color blind people (27)
Ellipse right: Color blind men (25)
Ellipse left: Color blind women (2)
Question: If you randomly selected 10 people from all the color-blind people in this town, how many would you expect to be men?

<u>Text RC3:</u>

T1: A small town has a population of town of 1,000 people (500 men and 500 women). A total of 27 people in the town are color blind.
Question A: If you randomly selected 100 people from this population, how many would you expect to be color blind?
T2: Of the color-blind people in the town, 25 are men.
Question B: If you randomly selected 10 people from all the color blind people in this town, how many would you expect to be men?

<u>Diagram RC3:</u>

T1: A small town has a population of both women and men, some of whom are color blind.
T2: The population can be depicted as follows:
Circle A label: Population of men and women (1000)
Circle B label: Color blind people (27)
Question A: If you randomly selected 100 people from this population, how many would you expect to be color blind?
Circle C label: Color-blind men (25)
Question B: If you randomly selected 10 people from all the color blind people in this town, how many would you expect to be men?

PROBLEM 7. SPAM E-MAIL

Prevalence: 48% (24/50)
Sensitivity: 83% (20/24)
Specificity: 77% (20/26)

H = Email is spam Not H = Email is not spam
D = Email flagged as spam Not D = Email flagged as clean

	H	Not H	
D	20	6	26
Not D	4	20	24
	24	26	50

Answer final: 20 / 26 = 76.92% * 200 = 153.84 = 154 or 153
Answer A: 26 / 50 = 52% * 1500 = 780

Text RC6:

T1: The ABC Technology Company has implemented spam filtering software. The company did a test on 50 messages. The test results on 50 messages showed that the software flagged 26 messages as spam and 24 flagged as clean. After further investigation into the fifty messages, the company found a total of 24 messages actually as spam, with 20 of those messages as correctly flagged as spam and 4 that were incorrectly flagged as clean.
Question: If the software flagged 200 new emails as spam, how many of those email messages would actually be spam?

Diagram RC6:

T1: The ABC Technology company has implemented spam filtering software.
T2: The company recently tested the software on some email with the following results:
Right circle label: Emails flagged as spam (26)
Left circle label: Email flagged as clean (24)
Ellipse: Email with spam (24)
Ellipse right: Emails flagged as spam actually spam (20)
Ellipse left: Emails flagged as clean actually spam (4)
Question: If the software flagged 200 new emails as spam, how many of those messages would actually be spam?

Text RC3:

T1: The ABC Technology Company has implemented spam filtering software. The company did a test on 50 messages. The test results on 50 messages showed that the software flagged 26 messages as spam.
Question A: If the software scanned another 1,500 messages, how many would be flagged as spam?
T2: Of the emails flagged as spam, 20 were actually spam.

Question B: If the software flagged 200 new emails as spam, how many of those messages would actually be spam?

Diagram RC3:

T1: The ABC Technology company has implemented spam filtering software.
T2: The company recently tested the software on some email with the following results:
Circle A label: Tested messages (50)
Circle B label: Emails flagged as spam (26)
Question A: If the software scanned another 1,500 messages, how many would be flagged as spam?
Circle C label: Emails flagged as spam actually spam (20)
Question B: If the software flagged 200 new emails as spam, how many of those messages would actually be spam?

PROBLEM 8. ARSON TRIAL

Prevalence: 11% (108/1000)
Sensitivity: 78% (84/108)
Specificity: 96% (856/892)

H = Person is actually a match   Not H = Person is not a match
D = Test indicates person is a match   Not D = Test indicates person is not a match

	H	Not H	
D	84	36	119
Not D	24	856	881
	108	892	1000

Answer final: 84 / 119 = 70.588% * 7000 = 4940, 4,941, 4942, 4935, 4900
Answer A: 119 / 1000 = 11.9% * 20,000 = 2380 (2200)

Text RC6:

T1: During an arson trial, a pathologist testified that the fingerprints found on the bottle of lighter fluid matched the suspect. But the pathologist testified that the particular test used to determine a match is not 100% accurate. If a sample of 1,000 residents were tested, 119 would have tests indicating they were a fingerprint match and 881 would have tests indicating no fingerprint match. However, because of test inaccuracy, the test would work correctly for only 84 of those 119 residents and would fail for 36 of those 119 residents. Overall, 120 residents would actually be a match.

Question: If you assume that 7,000 people have tests indicating a match, how many would you expect to be actually a match?

Diagram RC6:

T1: During an arson trial, a pathologist testified that the fingerprints found on the bottle of lighter fluid matched the suspect. But the pathologist testified that the particular test used to determine a match is not 100% accurate.
T2: The accuracy of the test on a sample of people is as follows:
Right circle label: People with tests indicating a match (119)
Left circle label: People with tests indicating no match (881)
Ellipse: People actually a match (120)
Ellipse right: People with tests indicating a match who actually are a match (84)
Ellipse left: People with tests indicating no match who actually are a match (36)
Question: If you assume 7,000 people have tests indicating a match, how many would you expect to be actually a match?

<u>Text RC3:</u>

T1: During an arson trial, a pathologist testified that the fingerprints found on the bottle of lighter fluid matched the suspect. But the pathologist testified that the particular test used to determine a match is not 100% accurate. If a sample of 1,000 residents were tested, 119 would have tests indicating they were a fingerprint match.
Question A: Based on this data, if a sample of 20,000 residents, how many do you expect will have a test indicating a match?
T2: Of those with tests indicating they were a fingerprint match, 84 people were later shown to be actually a match.
Question B: If you assume 7,000 people have tests indicating a match, how many would you expect to be actually a match?

<u>Diagram RC3:</u>

T1: During an arson trial, a pathologist testified that the fingerprints found on the bottle of lighter fluid matched the suspect. But the pathologist testified that the particular test used to determine a match is not 100% accurate.
T2: The accuracy of the test on a sample of people is as follows:
Circle A label: People tested (1000)
Circle B label: People with tests indicating a match (119)
Question A: Based on this data, in a sample of 20,000 residents, how many do you expect will have a test indicating a match?
Circle C label: People actually a match (84)
Question B: If you assume 7,000 people have tests indicating a match, how many would you expect to be actually a match?

PROBLEM 9. COMPROMISED PCS

Prevalence: 4% (15/400)
Sensitivity: 93% (14/15)
Specificity: 89% (343/385)

H = PC has been compromised　　Not H = PC has not been compromised
D = PC has possible vulnerability　　Not D = Does not have a vulnerability

	H	Not H	
D	14	42	56
Not D	1	343	344
	15	385	400

Answer final: 14 / 56 = 25% * 150 = 37.5 (37 or 38)
Answer A: 56 / 400 = 14% * 1200 = 168

<u>Text RC6:</u>

T1: A medium sized firm scanned a sample of its PCs and found some of its computers had vulnerabilities that might let hackers control the PCs. The scan also showed that some of its computers were already taken over by hackers. In this test, the security team scanned a sample of 400 PCs and found that 56 PCs out of the 400 were showing a vulnerability and 344 were not showing a vulnerability. Upon further examination by the security team, only 15 PCs were hacked into, with 14 that were hacked into as showing a vulnerability and 1 that was not showing a vulnerability.
Question: Based on this data, if the security team had another 150 PCs showing a vulnerability, how many would have been hacked into?

<u>Diagram RC6:</u>

T1: A medium-sized firm scanned a sample of its PCs and found some of its computers had vulnerabilities that might let hackers control the PC.
T2. The scan also showed that and some of its computers were already taken over by hackers. The details of the scan are as follows:
Right circle label: PCs showing a vulnerability (56)
Left circle label: PCs not showing a vulnerability (344)
Ellipse: PCs hacked into (15)
Ellipse right: PCs showing a vulnerability hacked into (14)
Ellipse left: PCs not showing a vulnerability hacked into (1)
Question: Based on this data, if the security team had another 150 PCs showing a vulnerability, how many would be hacked into?

<u>Text RC3:</u>

T1: A medium sized firm scanned a sample of its PCs and found some of its computers had vulnerabilities that might let hackers control the PCs. The scan also showed that some of its computers were already taken over by hackers. In this test, the security team scanned a sample of 400 PCs and found that 56 PCs were showing a vulnerability.
Question A: If the security team scanned another 1,200 PCs, how many would show a vulnerability?
T2. Of the PCs showing a vulnerability, 14 were hacked into.
Question B: If the security team had another 150 PCs showing a vulnerability, how many would be hacked into?

<u>Diagram RC3:</u>

T1: A medium-sized firm scanned a sample of its PCs and found some of its computers had vulnerabilities that might let hackers control the PC.

T2. The scan also showed that and some of its computers that were already taken over by hackers. The details of the scan are as follows:
Circle A label: PCs scanned (400)
Circle B label: PCs showing a vulnerability (56)
Question A: If the security team scanned another 1,200 PCs, how many would show a vulnerability?
Circle C label: PCs showing a vulnerability that were hacked into (14)
Question B: If the security team had another 150 PCs showing a vulnerability, how many would be hacked into?

PROBLEM 10. MARKETING STUDY

Prevalence: 60% (120/200)
Sensitivity: 40% (48/120)
Specificity: 68% (54/80)

H = Customer bought
D = Customer was thought likely to buy
Not H = Customer did not buy
Not D = Customer was thought unlikely to buy

	H	Not H	
D	48	26	74
Not D	72	54	126
	120	80	200

Answer: 48 / 74 = 64.864% * 10,000= 6,486.4 (6,486 or 6,487 or 6480, or 6400, or 6500)
Answer A: 74 / 200 = 37% * 50,000 = 18,500

<u>Text RC6:</u>

T1: A company did a study of its customers to predict who would be likely to buy the company's new product. As part of the study, the company sent an offer to buy the new product to 200 of its customers. The study identified 74 of the 200 customers as likely to buy the new product and 126 not likely to buy the new product. Of the 200 customers, a total of 120 customers bought the new product. Of those 120 customers who bought, 48 customers were identified as likely to buy, and 72 were identified as not likely to buy.
Question: If the company were to send the offer to buy the new product to another list of 10,000 customers who were identified as likely to buy the new product, how many would you expect to actually buy the new product?

<u>Diagram RC6:</u>

T1: A company did a study of some of its customers to predict who would be likely to buy the company's new product. As part of the study, the company sent an offer to buy the new product to some of its customers.
T2. The results of the mailing of the offer showed the following results:
Right circle label: Customers not likely to buy the new product (126)
Left circle label: Customers likely to buy the new product (74)
Ellipse: Customers who bought the new product (120)
Ellipse right: Customers not likely to buy who bought the new product (72)

Ellipse left: Customers likely to buy who bought the new product (48)
Question: If the company were to send an offer to a new list of 10,000 customers who were already identified as likely to buy, how many would you expect to actually buy the new product?

<u>Text RC3:</u>

T1: A company did a study of its customers to predict who would be likely to buy the company's new product. As part of the study, the company sent an offer to buy the new product to 200 of its customers. The study identified 74 of the 200 customers as likely to buy the new product.
Question A: If the company were to send an offer to buy the new product to a new list of 50,000 potential customers, how many would be likely to buy the new product?
T2: Of the customers who would be likely to buy, 48 bought the new product.
Question B: If the company were to send an offer to a new list of 10,000 customers who were already identified as likely to buy, how many would you expect to actually buy the new product?

<u>Diagram RC3:</u>

T1: A company did a study of some of its customers to predict who would be likely to buy the company's new product. As part of the study, the company sent an offer to buy the new product to some of its customers.
T2. The results of the mailing of the offer showed the following results:
Circle A label: Customers tested (200)
Circle B label: Customers likely to buy (74)
Question A: If the company were to send an offer to buy the new product to a new list of 50,000 potential customers, how many would be likely to buy the new product?
Circle C label: Customers likely to buy who bought (48)
Question B: If the company were to send an offer to a new list of 10,000 customers who were already identified as likely to buy, how many would you expect to actually buy the new product?