OVERCOMING BIASES TO IMPROVE SEARCH SUFFICIENCY AND DECISION ACCURACY: THE EFFECTS OF DATA VISUALIZATION, INSTRUCTIONS AND ORDER

by

Heidi S. Kramer

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

The University of Utah

May 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Heidi S. Kramer**

has been approved by the following supervisory committee members:

Frank A Drews	, Chair	**2/7/2014** Date Approved
Sarah Creem-Regehr	, Member	**2/7/2014** Date Approved
Frances J. Friedrich	, Member	**2/7/2014** Date Approved
Jonathan Nebeker	, Member	**2/7/2014** Date Approved
Jeanine Stefanucci	, Member	**2/7/2014** Date Approved

and by **Carol Sansone** , Chair/Dean of

the Department/College/School of **Psychology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The present study explored data presentation and human cognition with the objective of improving electronic Decision Support Systems (DSS). Computers have been used as tools for decision support for over 60 years, with the intent to supplement or replace human cognition. However, electronic computing has failed to reliably replace human cognition in complex domains. The suboptimal properties of the data and complexities of the domain often require human interpretation and intervention. Human interpretation relies on experience, values, intuition, insight and learning; which can lead to shortcuts or heuristics. Heuristics in the correct context can be economical and effective in solving many problems. When heuristics fail the results are labeled as cognitive biases or errors. Biases all share the elements of structuring incorrect or inappropriate models or hypotheses and/or insufficient consideration of the data. Most biases can be linked to confirmation bias – which is manifested by searches for and consideration of only confirming data. De-biasing techniques share the concept of shifting cognitive processing from an automatic associative mode to a more deliberate, conscious rule-based mode. This study used a modified Wason 2-4-6 task that combined methods of, 1) increased salience through data visualization with 2) appealing to the rule-based system through task instructions. The results indicate that neither increased salience nor instructions ensure increased search sufficiency, efficiency or decision accuracy. However, this study provides insight into the perceived value of evidence and four potential limitations related to self-directed searches: 1) The selection of necessary disconfirming evidence cannot be assumed, regardless of the perceived value of disconfirming evidence. 2) The selection of sufficient evidence does not ensure accuracy; however, 3) insufficient selection of disconfirming evidence results in lower accuracy. 4) Ambiguous evidence is considered more valuable than potentially disconfirming evidence. Implications for the design of decision support systems are presented along with limitations and directions for future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Computers as Tools for Decision Support

Since the early days of electronic computing in the 1940s and 1950s, researchers have worked to replace or supplement human cognition with Decision Support Systems (DSS) and Natural Language Processing (NLP); (Dick, Steen, & Detmer, 1997; Liddy, 2001; Turban & Aronson, 2000). No universally accepted definition exists for either DSS or NLP; however, the term Decision Support System can be used to describe any electronic system that supports decision making (Turban & Aronson, 2000); and NLP refers to computerized approaches to analyzing text with the intent "to accomplish human-like language processing" (Liddy, 2001).

The success of DSS and NLP in complex environments is complicated by data characteristics such as conflicting, ambiguous, and excessive data with issues of availability, timeliness and credibility. The suboptimal properties of existing data are recognized in healthcare where even the definition of evidence-based medicine "makes allowances for *missing, incomplete*, or *low-quality evidence* [italics added]" (Sim et al., 2001).

Unmet Expectations

Researchers and developers continue the struggle to provide DSS systems that reliably replace human cognition in complex domains such as healthcare, weather forecasting and security intelligence. (MacEachren et al., 2005; Puvathingal & Hantula, 2012; Sittig et al., 2010; Wright et al., 2011). For over 60 years the focus has been on the capabilities of the electronic system. Perhaps the time has come to focus more on human cognition, which requires the development of electronic solutions structured to support human cognitive strengths and overcome cognitive deficiencies (Hollnagel & Woods, 2005).

Heuristics and Biases

Cognitive strengths can be associated with cognitive deficiencies. Experience, values, intuition, insight and learning lead to heuristics or cognitive shortcuts (Simon & Newell, 1958). When applied in the correct context, heuristics are economical, effective and are often the result of domain expertise; many real world problems are effectively solved using heuristics (Gigerenzer, 2007; Klayman & Ha, 1987). However, when heuristics fail they are labeled as cognitive biases or errors, using terms such as, gambler's fallacy, premature cognitive closure, representativeness, availability, and confirmation biases (Croskerry, 2002; Simon, 1956; Tversky & Kahneman, 1974). Whatever their label, biases all share the elements of, a) structuring incorrect or inappropriate models and hypotheses and/or, b) insufficient consideration of the data. Using Nickerson's (1998) definition of confirmation bias as the *"unwitting selectivity in the acquisition and use of evidence…*[italics added]*"* it appears most biases can be linked to *confirmation bias*. Confirmation bias manifests itself by searches for only confirming information, that are insufficient and lead to lower decision accuracy (Nickerson, 1998; Wason, 1960). Poor outcomes due to incorrect models or hypotheses and or insufficient consideration of data are ubiquitous. Many researchers report poor outcomes from deficient decisions and discuss the need to overcome confirmation bias; however, few propose de-biasing techniques (Booth, Carroll, Ilott, Low, & Cooper, 2013; Lilienfeld, Ammirati, & Landfield, 2009). The debiasing techniques that do exist share the concept of shifting cognitive processing from an automatic, associative mode to a more deliberate, conscious rulebased mode (Lilienfeld et al., 2009).

Dual Processing

Two Systems

At least since Aristotle, human cognition has been thought of as two systems or to consist of dual processing: An automatic, unconscious, fast system, which includes heuristics; and a rule-based, conscious, deliberative system (Evans, 2008; Kahneman, 2011; Sloman, 1996). The use of dual processing systems is not limited to any domain, problem or context and depends on the "reasoner's knowledge, skill and experience" (Sloman, 1996). Multiple definitions and descriptions in dual process theory have resulted in many labels for the two systems; including, Automatic and Controlled, Heuristic/Intuitive and Systematic/Analytic, Associative and Rule-based; or simply System 1 and System 2 or Type 1 and Type 2 (see Evans, 2008, p. 257).

Researchers typically agree the two systems work in parallel with competition for controlling behavior and decisions (Barrouillet, 2011; Evans, 2003; Kahneman, 2011; Sloman, 1996; Stanovich, West, & Toplak, 2011). However, the diverse descriptions and development of theories of the two systems continue and many unknowns remain when defining neural correlates (Barrouillet, 2011; Evans, 2011; Tsujii & Watanabe, 2009). For consistency, the two systems are referenced hereafter using Sloman's (1996) terms of "associative" and "rule-based." These terms are more descriptive than System or Type 1 and 2 and are at least as explanatory as labels such as Automatic and Controlled, Heuristic/Intuitive and Systematic/Analytic.

One objective of the present research was to build on the concept that shifting cognitive processing from an automatic, associative mode to a more deliberate, conscious rule-based mode may be crucial to overcoming confirmation bias. Therefore, the question became when and how to facilitate that shift in the context of decision support.

Switching Between Systems

To improve human cognition by reducing confirmation bias, it may be important to understand when to support associative reasoning and when and how to facilitate switching to more rule-based processing. There are at least two broad approaches to supporting associative reasoning and facilitating a switch to rule-based processing: 1) Data visualizations to provide effective and efficient explanations of the data (Kosslyn, 1989; Tufte, 2000) and to increase awareness of data characteristics; and, 2) training or explicit instruction to recognize errors, structure appropriate hypotheses, and/or conduct deliberate, methodical searches and analyses. These two approaches are discussed in detail.

Data Visualization

Data visualizations are often intended to provide effective and efficient explanations and support decision making. Data visualizations are typically thought of as the use of charts and graphs that show data trends and groupings. These data visualizations can be effective; however, they often do not include representing characteristics of the data such as, temporal qualities, precision, assumed accuracy, data source credibility, measurement reliability, availability and the relationship of the data to the current task. An overview of work promoting data visualization revealed a common focus on a limited set of principles and techniques: How relevant data can be made salient; warnings that graphical displays must be portrayed so that comparisons can be accurate and not mislead; and simplicity for easy comprehension (see Few, 2006; Kosslyn, 1989; Tufte, 2000; Wainer, 2005). Although this overview was not exhaustive, it is surprising to find that the authors never addressed and only rarely acknowledged issues with data characteristics that affect the user's perception of data quality, ambiguity or uncertainty as important elements. One exception is a book on graphing history and principles where the author acknowledged, "The focus of this book has been on displaying data, I have spent very little effort discussing the quality of what is being displayed" (Wainer, 2005, p. 142) This void is notable since data quality, ambiguity and uncertainty can be important to determine the relevance of the data to the user's task (Lapinski, 2009; Puvathingal & Hantula, 2012). Only recently have some researchers started to consider how to represent data characteristics that affect quality, uncertainty and ambiguity (MacEachren et al., 2005; Pang, Wittenbrink, & Lodha, 1997; Zuk & Carpendale, 2006). Consequentially, there is very little understanding of the effect on human cognition of representing data characteristics.

Zuk and Carpendale (2007) suggest that increasing awareness of data characteristics may lead to consideration of more causes and reduce biases. Awareness of data characteristics may be raised by increasing the *salience* of data characteristics, especially those characteristics that are relevant to the current hypothesis or task goals. Salience has been defined as how perceptually distinctive the information is in the environment in which it is presented (Fishbein, Haygood, & Frieson, 1970; Sanfey &

Hastie, 1998). This perceptual distinctiveness is achieved by stimulus-driven or bottom-up processing which is based on display properties such as size, position, line orientation, sound levels, stimulus-onset, or color hue and intensity; all of which can affect the search for information (Huang & Pashler, 2005; Nothdurft, 1993; Treisman & Gelade, 1980; Treisman, Gelade, & Yantis, 2000).

Natural Language Processing (NLP) tools have been created to electronically annotate text to increase the salience of some data characteristics (see Chen & Styler., 2013; South et al., 2012). In healthcare, electronically annotating or tagging evidence may help identify patients with certain medical conditions or potentially diagnose a condition in a single patient. Evidence may be annotated in any number of ways; including highlighting text by, for example, using different colors to identify and categorize data as supporting, ambiguous or disconfirming in relation to a specific medical condition. More specifically, consider the medical diagnosis of pneumonia; the evidence in a patient's records such as the symptoms of cough, shortness of breath, or chest pains, may be annotated as supporting a pneumonia diagnosis while other evidence such as nausea, vomiting or diarrhea may be annotated as ambiguous. A clear chest x-ray can be annotated as disconfirming the diagnosis of pneumonia. Using computer software to find and annotate evidence as supporting, ambiguous or disconfirming (using colors or other methods to differentiate categories and increase salience) may lead to more efficient and sufficient searches with greater decision accuracy. Thus, the question becomes, can increased salience of data characteristics reduce confirmation bias and thereby improve performance; perhaps by appropriately supporting associative processing or facilitating a switch to rule-based reasoning.

Instructions

A second approach to reducing confirmation bias is to encourage rule-based processing with instructions to recognize errors, structure appropriate hypotheses, and/or conduct deliberate, methodical searches (Edward R. Hirt, Kardes, & Markman, 2004; Edward R Hirt & Markman, 1995; Lilienfeld et al., 2009; Spengler, Strohmer, Dixon, & Shivy, 1995). As noted earlier, structuring incorrect models and hypotheses, and insufficient consideration of the data are core elements and failure points of biases. These failure points are not isolated or separate and distinct processes; without an accurate hypothesis the value of the data is not interpreted correctly and therefore the wrong data are considered or the correct data are not given sufficient or accurate consideration (Wason & Johnson-Laird, 1970). Two types of instructions for overcoming confirmation bias have been considered, creation of multiple hypotheses and delaying judgment while considering more data – especially disconfirming data (see Lilienfeld et al., 2009).

Many practices have been proposed for encouraging creation of multiple hypotheses, such as, “consider-the-opposite” (Lord, Lepper, & Preston, 1984) and later Hirt and Markham’s (1995) proposal to consider any possible outcome, not just the opposite strategies. In medicine considering any possible outcome is supported by the concept of *differential diagnosis*. Differential diagnosis is related to a “hypothetico-deductive method” where doctors are encouraged to iteratively consider evidence using bottom up deductive processes to develop hypotheses then seek more evidence (e.g., laboratory and radiology tests), to make a diagnosis (Croskerry, 2002).

The second type of instruction to reduce confirmation bias – delaying judgment while considering more data – may also appeal to the deliberate rule-based system of reasoning. Clinical counselors who delay judgment in diagnosis have been shown to be more accurate; perhaps due to forming hypotheses that are "repeatedly subjected to rejection" based on additional evidence (Spengler et al., 1995). The question is, can instructions to structure appropriate hypotheses and sufficiently consider data, reduce confirmation bias and thereby improve performance?

Overview of Experiments

The present research included three experiments. All experiments explored means to reduce or to overcome cognitive biases by promoting, a) the generation of appropriate models/hypotheses and, b) sufficient consideration of the data. The first experimental design combined increased salience of data characteristics with instructions. This study addressed the following two questions: Can increased salience of data characteristics reduce confirmation bias and thereby improve performance; and, can instructions to structure appropriate hypotheses and consider data sufficiently, reduce confirmation bias and thereby improve performance?

The objectives of the second experiment were to delve deeper into the perceived value and consideration of data; and continue exploration of increased salience of data characteristics. The third experiment removed the ability for self-directed searches by presenting the data in a specific order, and did not present ambiguous data. The effect of presenting sufficient disconfirming evidence was explored in the analyses of all experiment conditions; specifically considering the following question: How does the presence of disconfirming evidence affect cognition.

GENERAL METHOD

Wason 2-4-6 Task

All approaches in the present research were implemented using adaptations of the Wason 2-4-6 task. Variations of the Wason 2-4-6 task have been used extensively to study hypothesis generation, and confirming and disconfirming search strategies (see Cherubini, Castelvecchio, & Cherubini, 2005; Gorman & Gorman, 1984; Mumma & Wilson, 1995). The original Wason 2-4-6 task was to discover the rule governing the sequence of three numbers – a triple. Participants were presented with a triple and asked to discover the rule that described the relationship between the three numbers. Next, the participants proposed triples to test their hypotheses on what the rule could be; then received feedback on whether or not the proposed triple adhered to the rule describing the original triple. The participants generated triples until they were confident they had discovered the rule; they then stated the rule and received feedback on whether their rule was correct. If incorrect, the participant formulated another hypothesis for the rule and generated more triples. This process continued until the participants discovered the rule, abandoned the task or until time expired. The original Wason 2-4-6 task presented the triple “2-4-6” where the rule was “three numbers in increasing order of magnitude” (Wason, 1960).

The Wason 2-4-6 task seems simple; however, it is rather difficult. In Wason’s original study only 6 of the 29 participants (21%) discovered the rule on their first declaration of a rule; another 10 participants (34%) discovered the rule on their second attempt. Eight participants (28%) never discovered the rule.

The central element of a successful discovery of the rule is to seek disconfirming evidence. Therefore, the process of rule generation in a variation of the Wason 2-4-6 task is well-suited to study methods for overcoming confirmation bias; since triples can be definitively labeled as conforming, nonconforming, or partially conforming to a rule. In addition, the Wason 2-4-6 task provides the benefits of unambiguous measures for search strategies, sufficiency, and efficiency and decision accuracy.

General Design

All three experiments were similar in that, consistent with the original Wason 2-4-6 task, participants were asked to discover a rule related to sequences of numbers (triples); (see Figure 1). However, unlike the original Wason 2-4-6 task, instead of a single trial with one triple and associated rule, participants were given 15 trials; each relating to a different rule (see Table 1). In addition, the method used here also differed from the classic Wason 2-4-6 task in that the participants did not actively generate triples; instead, participants selected or viewed up to 4 or 6 triples for each trial. Some of the triples conformed to the rule, some partially conformed to the rule and other triples did not conform to the rule. These data characteristics of conforming, partially conforming and not conforming are analogous to evidence being confirming, ambiguous or disconfirming a hypothesis or diagnosis. After selecting and considering the triples participants then stated what they thought the rule could be.

The structure of the rules was consistent across all three experiments. The rules were built on three pattern types: 1) The numbers being odd or even; 2) the difference

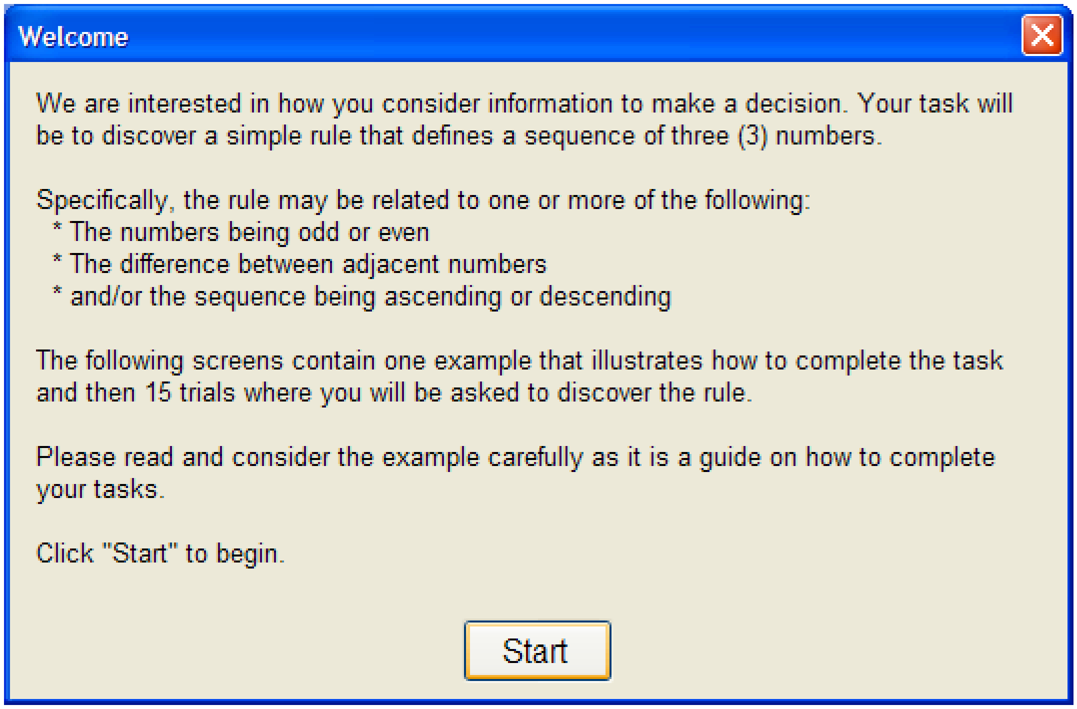

Figure 1. Welcome screen

Table 1. Rules for the 15 Trials

Trial	Rule
1	A difference of 1 must separate adjacent numbers.
2	A difference of 0 or 1 must separate adjacent numbers.
3	Numbers must alternate odd/even.
4	Ascending odd numbers.
5	Odd numbers where a difference of 2 must separate adjacent numbers.
6	Ascending numbers where a difference of 3 must separate adjacent numbers.
7	Descending odd numbers.
8	Any even number.
9	Ascending numbers which must alternate odd/even.
10	Descending numbers where a difference of 2 must separate adjacent numbers.
11	Descending numbers.
12	Even numbers where a difference of 4 must separate adjacent numbers.
13	Even numbers where a difference of 2 or 6 must separate adjacent numbers.
14	Any odd number.
15	Descending numbers where a difference of 2 or 3 must separate adjacent numbers.

between adjacent numbers; and/or 3) the sequence being ascending or descending. For example, the triple 6-4-2 would conform to the rule "even numbers descending by 2;" which includes all three patterns. In the instructions, the participants were cautioned that the conforming sequences could include patterns that were not required for the rule; and to limit the rule to only the required patterns. For example, the rule for the triple 6-4-2 could be "any even number." Determining what patterns to include in or exclude from the rule required consideration of other conforming and nonconforming triples. The rule was given to the participant after every trial (see Figure 2). Participants completed the trials individually.

After completing the 15 trials, demographic and academic data were collected for each participant (see Figure 3): Gender, age, GPA and whether they had completed, or were currently enrolled in specific psychology classes (i.e., Psych 1010 – General Psychology, Psych 2010 – Psych as a Science and Profession, Psych 2125 – Everyday Decision Making, Psych 3120 – Cognitive Psychology, Psych 3171 – Human Factors). Participation in these courses was of interest because the courses may have included discussions of the Wason 2-4-6 task and thus provided some training to consider disconfirming evidence.

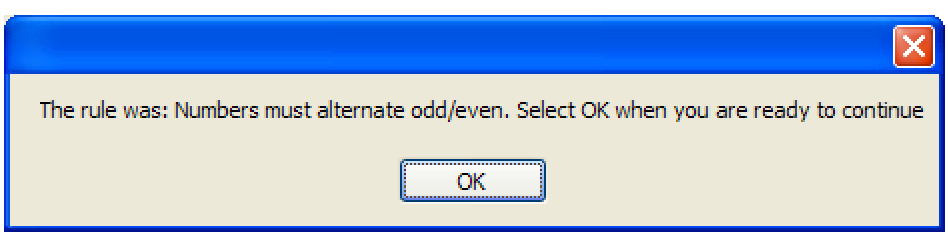

Figure 2. Providing the correct rule for the trial

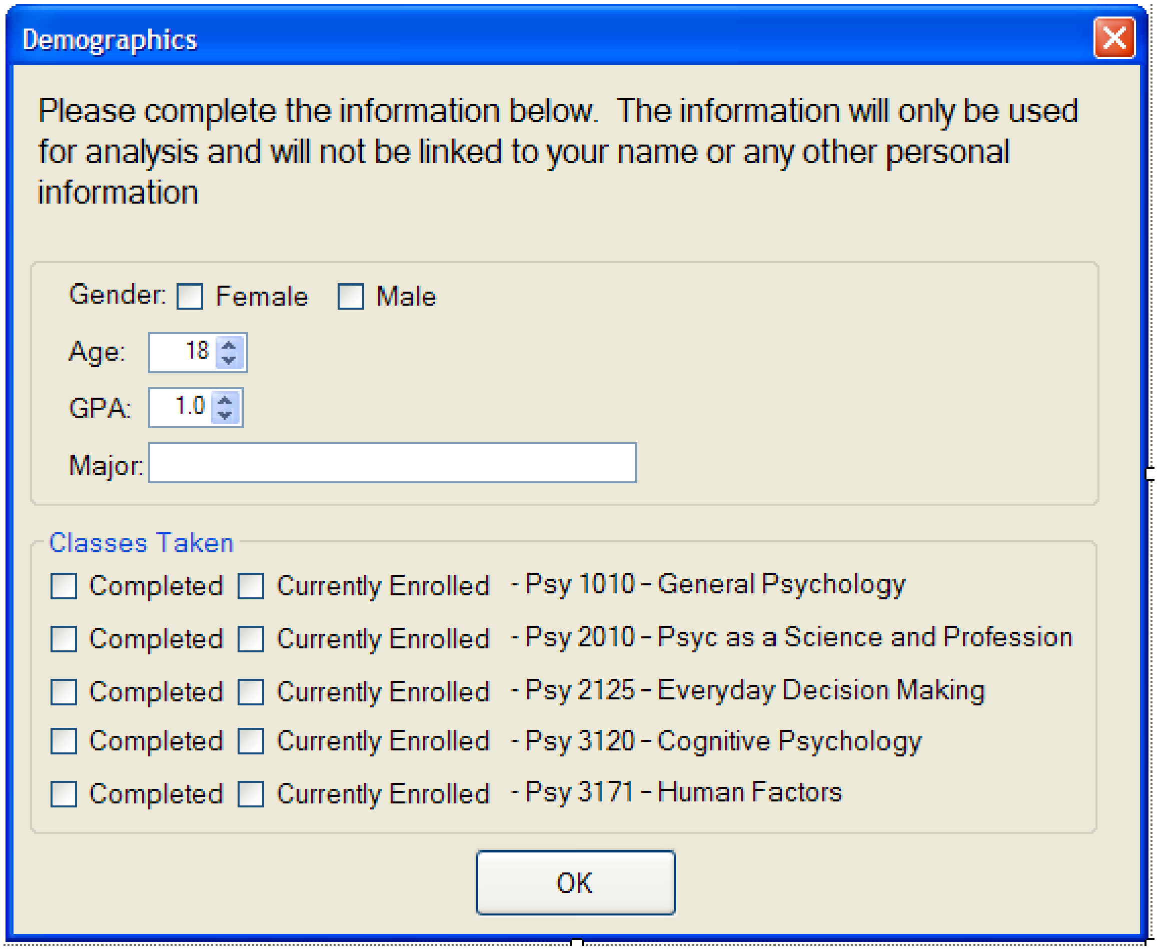

Figure 3. Demographic and academic information collection

General Data Coding

Rules

Research assistants coded the participants' responses for analysis. There were two parts to coding the rules, "hits" and "false positives." When participants correctly identified a pattern in the rule it was counted as a "hit." Rules consisted of one or two patterns: Six rules had only 1 pattern; nine rules had 2 patterns (see Table 1). To normalize identifying patterns in the rule to a maximum value of "1" per trial, the number of hits was divided by the total patterns in that rule. When participants included a pattern in the rule that should have been eliminated (a distracter pattern) the pattern was counted as a "false positive."

Because participants entered free form text when defining the rule, the participants' descriptions of the rules varied. To ensure reliability of coding, each participant's responses were independently coded by at least two research assistants. All coding was reviewed independently and differences between coders' counts of hits and false positives were investigated and resolved for consistency. For example, the rule for trial 12 was "Even numbers where a difference of 4 must separate adjacent numbers." Participant 231 entered "even numbers ascending must have difference of 4 between adjacent numbers." There were three possible patterns types (odd or even numbers, ascending or descending, the difference between the numbers) that could be counted as hits or false positives. For this example having "even numbers" and "difference of 4" were counted as two "hits" while the "ascending" pattern was counted as a "false positive." While coding one RA could have missed the false positive or only counted one hit, while another RA coded the participant's attempt correctly. Discrepancies between RA coding were identified and fixed.

Selection Sequence

The selection sequence was also coded for each trial per participant. For example, CCPPNN equates to a selection pattern of selecting both conforming triples (CC), then both partially conforming triples (PP) and finally both nonconforming triples (NN). Because coding the selection sequence was based on the data captured by the software the sequences were coded only once by a single RA and then checked for accuracy.

EXPERIMENT 1 - INCREASED SALIENCE AND TASK INSTRUCTIONS

Introduction

The first experiment combined increased salience of data characteristics with task instructions. The experiment addressed two potentially complementary questions. The first question was, can increased salience of data characteristics reduce confirmation bias and thereby improve performance? More specifically, with the hypothesis that color can be more effective than text in identifying characteristics of evidence, and therefore, replace text for increased search efficiency, sufficiency and decision accuracy.

Data visualization may not be enough to ensure sufficient consideration of the data. The second question for this experiment was, can instructions to structure appropriate hypotheses and sufficiently consider data, reduce confirmation bias and thereby improve performance? The two types of instructions for overcoming confirmation bias were introduced earlier – creation of multiple hypotheses and delaying judgment while considering more data, especially disconfirming data. Both types of instructions implicitly address structuring hypotheses and sufficient consideration of the data. The emphasis in past research has been on structuring the hypothesis, and secondarily on searching for disconfirming data. The search for disconfirming data is necessary for considering disconfirming data. Therefore, instructions to search for and consider disconfirming data may improve accuracy more than instructions to structure multiple hypotheses. The hypothesis for this experiment was that instructions to perform a disconfirming search are more effective than instructions to think of as many rules as possible. Compared to an emphasis on structuring hypotheses, an explicit emphasis on searching for disconfirming evidence might promote a switch to rule-based processing and increase sufficient consideration of the data (measured by selection of confirming and disconfirming evidence). Sufficient consideration of the data was expected to result in increased decision accuracy.

Method

Participants

Participants were recruited from the University of Utah psychology undergraduate student participant pool. Each participant individually completed one experiment in one session. Participants received one hour of credit towards a psychology course requirement. Data were analyzed from 80 participants ranging in age from 18 to 43 years ($M = 23$). Forty-eight participants were female (60%) and 32 were male (40%).

Exclusions

All potential participants were tested for color vision and their data were excluded from analyses if they could not distinguish between the colors used to identify data characteristics. In addition, 4 potential participants did not provide a rule for every trial; therefore, their data were incomplete and excluded from analysis.

Design

The 2x2 between-subjects design included visualization (text labels or color coding) as one independent variable and instructions (many rules or disconfirm) as the other independent variable. These variables are described in more detail in the following section.

Visualization

The visualization conditions were used to examine the following question: Can increased salience of data characteristics reduce confirmation bias and thereby improve performance, with the hypothesis that color can be more effective than text in identifying characteristics of evidence, and therefore, can replace text for increased search efficiency, sufficiency and decision accuracy.

Six boxes containing triples were presented in a column – triples were initially not visible. When selected by the participant each box revealed a triple. Each of the boxes for the triples were identified as “conforming,” “nonconforming,” or “partially conforming” to a rule using one of two methods: 1) Text labels or, 2) color coding – green, yellow and red – to match rule conformation (i.e., green = conforming; yellow = partially conforming; red = nonconforming); (see Figure 4 and Figure 5). The colors green, yellow and red were chosen because of their common associations: Green with go or yes; yellow with caution or maybe; and red with stop or no.

The boxes containing triples were positioned in pseudo-random order, with a different order for every trial but consistent between conditions for each trial (i.e., the positions in the list for conforming and nonconforming triples differed between trials 1 through 15; however, the order of conforming and nonconforming triples in trials 1 through 15 were the same for all participants); (see Table 2).

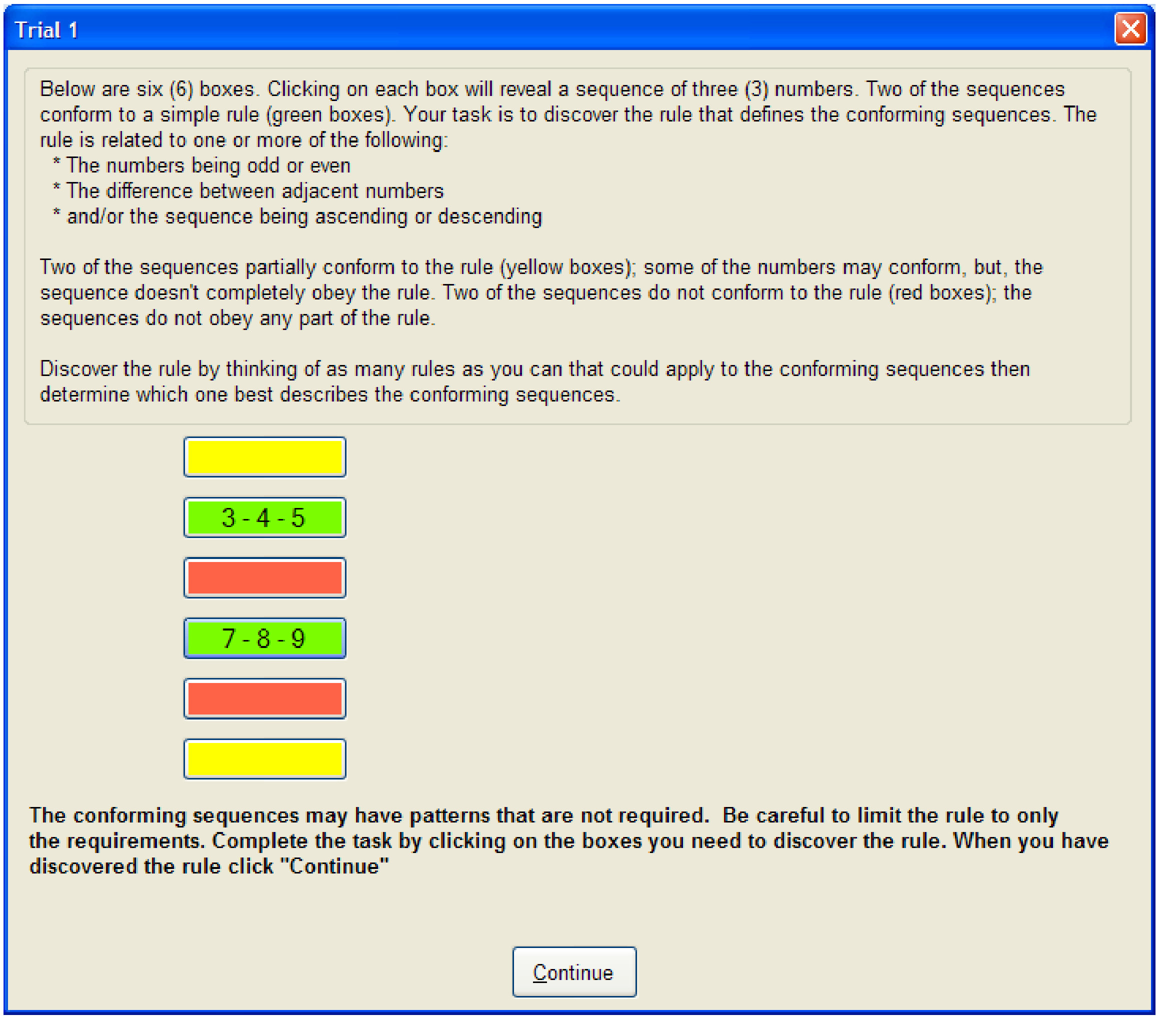

Figure 4. Color identification with “Many Rules” instructions

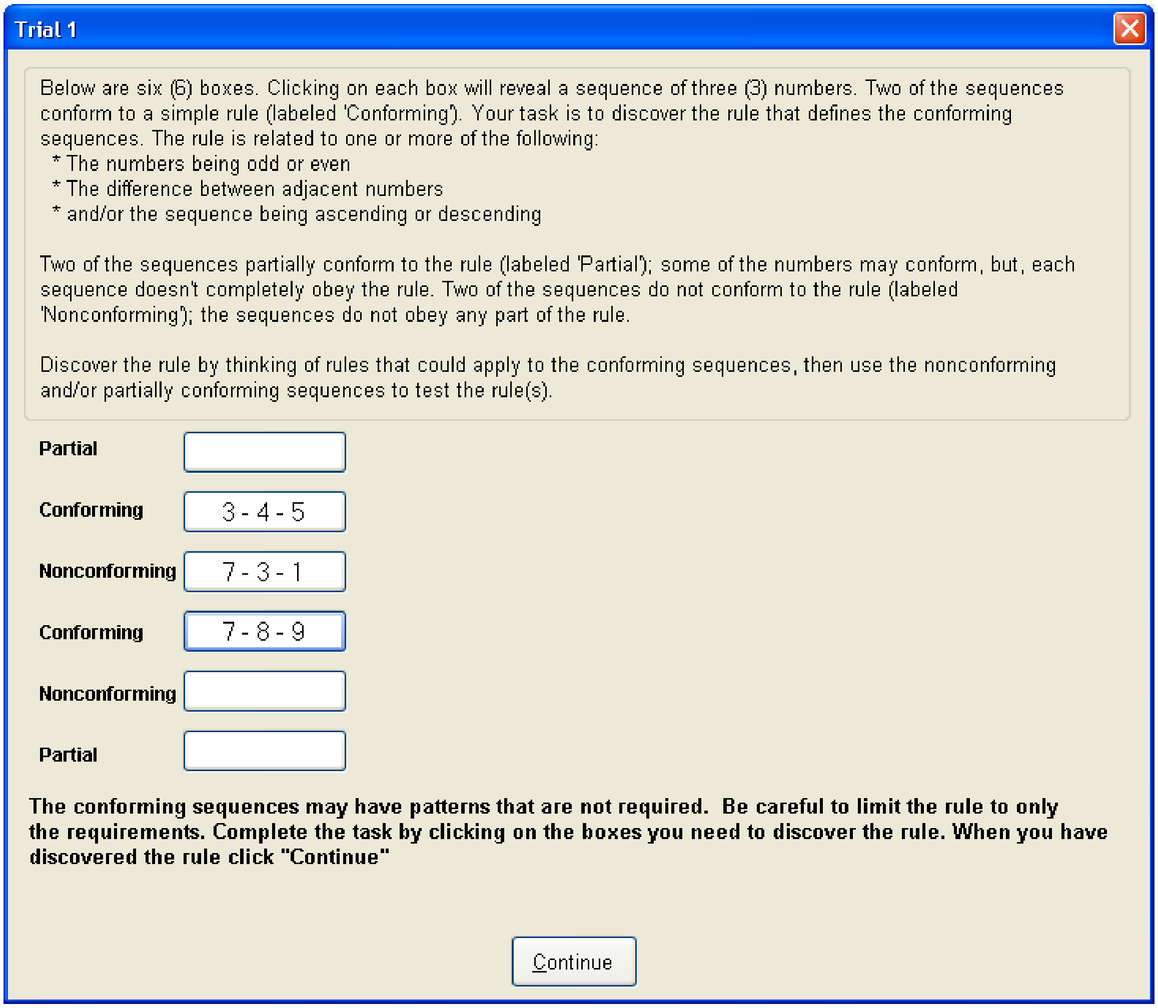

Figure 5. Text labels with “Disconfirming” instructions

Table 2. Experiments 1 and 2 - List Order for the 15 Trials

Trial	List Order
1	PCNCNP
2	NPPCNC
3	NPCNCP
4	CPNCNP
5	NNPCCP
6	PNNPCC
7	PCNPCN
8	PNCNCP
9	PCNCNP
10	CCPPNN
11	PNPCNC
12	CPNNCP
13	PPCNNC
14	NPCNPC
15	CPPNNC

Instructions

The use of different instructions explored the question, can instructions to structure appropriate hypotheses and sufficiently consider data reduce confirmation bias and thereby improve performance? The hypothesis was that instructions to perform a disconfirming search are more effective than instructions to think of as many rules possible. Expected results were increased search efficiency, sufficiency and decision accuracy.

Two versions of the instructions were provided in the between-subjects design. The first instructed participants to think of rules and then to perform a disconfirming search: “Discover the rule by thinking of rules that could apply to the conforming sequences then use the nonconforming and/or partially conforming sequences to test the rule(s).” The second instructed participants to generate as many rules as possible with no reference to disconfirming evidence: “Discover the rule by thinking of as many rules as you can that could apply to the conforming sequences then determine which one best describes the conforming sequences."

Procedure

Upon entry into the laboratory, participants were greeted and informed consent was obtained for the IRB approved study. Because the selection of triples potentially required discriminating between different colors, all potential participants were given a color vision screening test. Potential participants who could not discriminate between colors were assigned participant IDs that identified them as having color vision deficiencies and their data were excluded from all analyses. Participants completed the tasks individually. A custom software application introduced participants to the task, presented instructions, and recorded responses.

Using the software application, a research assistant (RA) logged in and entered a predetermined unique participant ID. The participant ID determined the participant's experimental condition. The RA directed the participant to read the instructions aloud. Participants were encouraged to ask questions while they completed an example trial. The example trial was used to train the participant on the task, address questions and assess the participant's understanding of the task. Participants generally understood the task; however, a common question was whether the rule could include more than one pattern. RAs were instructed to answer questions as "yes" or "no" or by rereading the relevant instructions on the computer display. A pilot of the study showed that rereading task instructions and pausing to allow participants to think about the task resolved any questions.

Participants interacted with the application by clicking on six initially empty boxes. The boxes were identified, either with text labels or by color, as containing triples that 1) conformed to the rule, 2) did not conform to the rule, or, 3) partially conformed to the rule (two triples of each type). After selecting a box a triple appeared in the box. Participants were able to select the boxes in any order and could select up to all six boxes. Text on the screen directed the participants to click "Continue" when they discovered the rule. Selecting the "Continue" button caused the boxes with the triples to disappear and displayed the text, "Write the rule you think defines the sequence. Click 'Continue' when you are finished." Participants entered the rule in a text box (see Figure 6). Selecting the "Continue" button caused the boxes and the triples that were selected to reappear and displayed the text, "Write the strategy you used to determine the rule. Click 'Continue' when you are finished." Participants entered the strategy they used to determine the rule. Selecting the "Continue" button caused the correct rule to appear (see Figure 2). After dismissing the dialog box containing the correct rule, the screen was reset to the next trial. After 15 trials the participants answered demographic questions (i.e., age, gender, GPA and classes taken); (see Figure 3).

Measurements

The following measurements were recorded in a database for each trial performed by a participant:

Time. The time was recorded for participant actions:

- When each triple was selected (made visible)
- When the participant determined the rule and selected "Continue").

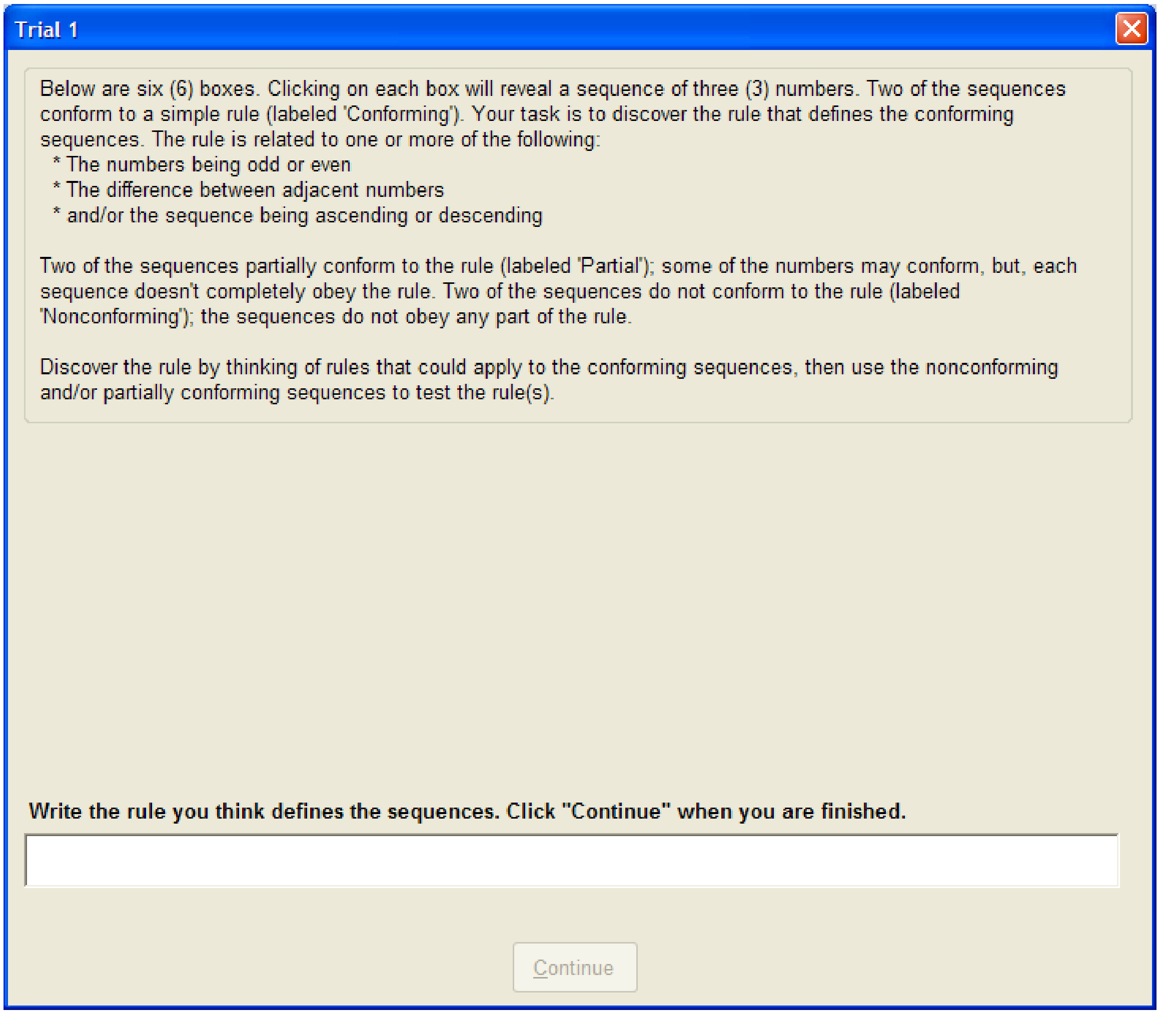

Figure 6. Prompt for the rule after selecting all triples

- When the participant selected “Continue” after entering the rule and the strategy for each trial (see Figure 7)

These measurements allowed comparisons of the time for trials between and within participants, and for each experimental condition. Search efficiency was measured by time.

Rule. Participants were instructed to enter what they thought the rule could be after selecting up to six triples. For every trial the conforming triples contained one or two patterns related to a rule and at least one other pattern that served as a distracter. The distracter pattern(s) could be eliminated from the rule using other triples (see Table 3).

Search sequence. When each triple was selected and made visible the type of sequence – conforming, nonconforming or partially conforming – was recorded as well as the time of selection. This facilitated determining the search sequence.

Strategy. The participant entered a description of the strategy they used to determine each rule. The example trial displayed the following text as a description of potential strategies (see Figure 7): “There are multiple strategies for determining the rule, one strategy is to look at just the first conforming sequence (3-5-7) and think of possible rules for that sequence, (e.g., the rule could be odd numbers with adjacent numbers differing by 2 and/or ascending numbers). Looking at the next conforming sequence tells you the rule can’t limit the numbers to odd only.” For the experimental conditions with the disconfirm instruction the text attempted to encourage the participant to disconfirm their hypothesized rule by continuing with, “You could hypothesize that the rule is ascending numbers differing by 2; then check the nonconforming and/or partially conforming sequences to support or contradict your hypothesis until you determine the

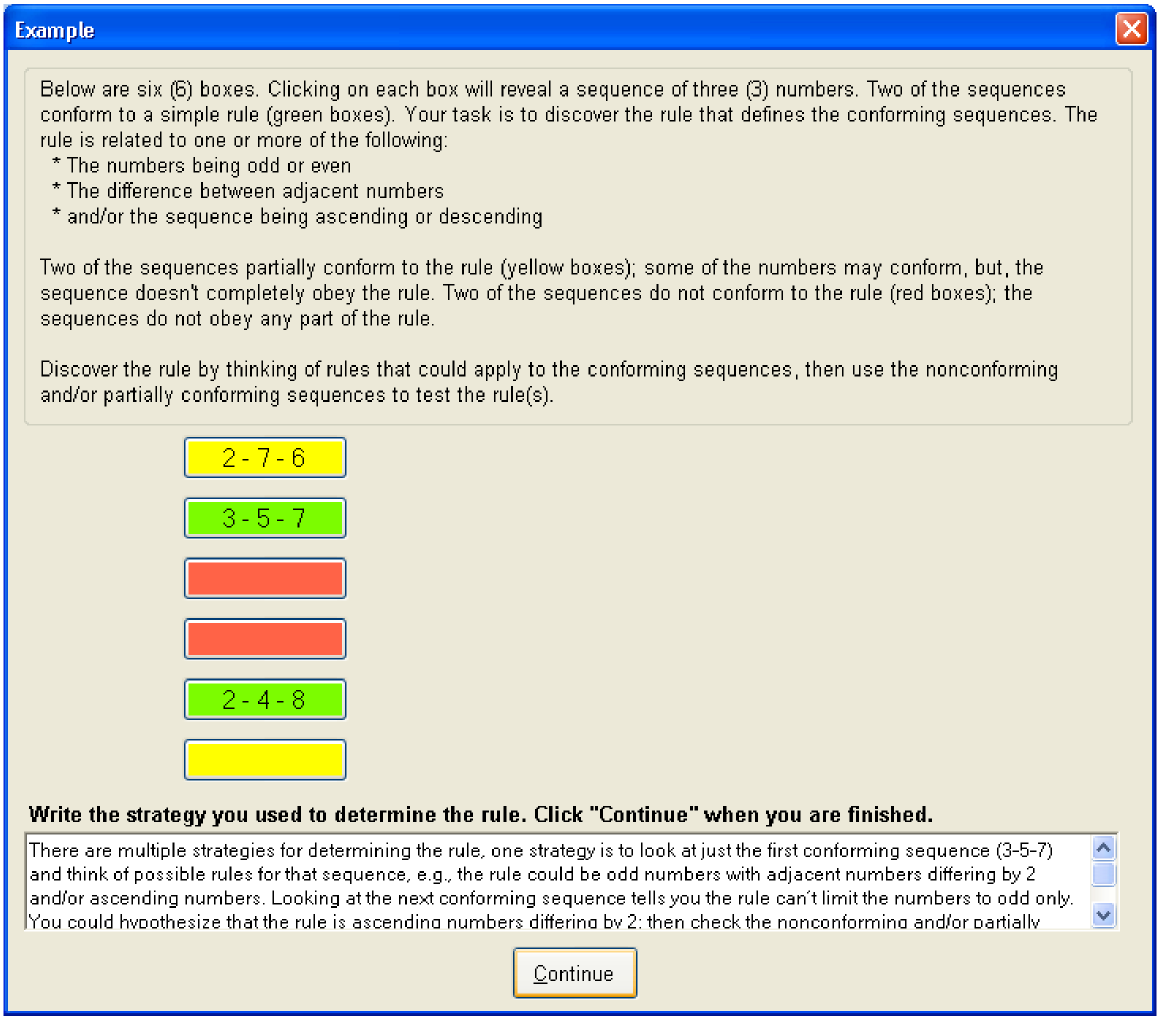

Figure 7. Example trial with explanatory text for inputting strategy

Table 3. Patterns in the Rule and Possible Distracters

Trial	Hit (part of rule)	Possible Hits	False Positive (distracter) – disconfirm with: Conforming	Nonconforming
1	Diff 1	1		Ascending-N2 ; Odd/even/odd-N2
2	Diff 0 or 1	1		Descending-N1
3	Odd/even/odd	1	Diff 1 - C2	Ascending-N1
4	Ascending, odd	2		Diff 2,4-N1
5	Diff 2, odd	2	Descending - C2	
6	Ascending, diff 3	2	Odd/even/odd - C2	Odd/even/odd-N1; Even/odd/even-N2
7	Descending, odd	2	Diff 4 then 2 - C2	Diff 4,2-N1 N2
8	Even	1	Descending - C1 Diff 4,2 - C1	
9	Ascending, even/odd/even	2	Diff 3,1 - C2	
10	Descending, diff 2	2	All odd - C2	All odd or even-N1 N2
11	Descending	1	Diff 2 - C2	Diff 2-N2; odd-N2
12	Diff 4, even	2	Ascending - C2	Ascending-N1
13	Diff 2 or 6, even	2		Descending-N2
14	Odd	1		Diff 2,4-N1; descending-N2
15	Descending, diff 2 or 3	2		Even/even/odd-N2

rule is just ascending numbers." The experimental conditions with the many rules instruction continued instead with, "You could hypothesize that the rule is simply ascending numbers." For all experimental conditions, the example strategy text concluded with, "Other strategies may be more or less effective; you are encouraged to develop your own strategies." The purpose of the example text was to describe what was meant by a "strategy" without specifying what the participant's strategy should be. The text box where participants typed their responses allowed sufficient input for all responses – no responses were truncated. Responses were saved as plain text.

Data Coding

The "hits" and "false positives" in the participant rule responses and the selection sequence were coded as described in the General Data Coding section.

Times

All times were calculated electronically using values captured and stored by the software application. Efficiency was judged by how much time the participant spent determining the rule. Time measurement for each trial started when the trial was presented and ended when the participant selected the "Continue" button to enter the rule.

Data Exclusion

Unfortunately, the text labels for trial 14 were incorrectly labeled in a version of the software application. Because this affected the responses for some participants, the data for trial 14 were excluded from analyses for all participants. The rule for trial 14 was unique (any odd number); however, other rules included the pattern of odd numbers.

Therefore, the trial 14 data were somewhat redundant and exclusion of trial 14 data was unlikely to have affected the overall results.

Results

Accuracy

A two-way ANOVA of the data showed no difference in accuracy for identifying patterns in the rule based on the use of text or color to identify data characteristics, $F(1, 76) = .22$, $p = .64$ (Text, $M = 9.5$ (67.9%), $SD = 1.85$; Color, $M = 9.3$ (66.4%), $SD = 2.22$). On average, participants failed to eliminate 6.58 distracter patterns ($SD = 2.68$) in the Text condition and 7.43 distracter patterns ($SD = 3.56$) in the Color condition. Resulting in no statistical difference in eliminating distracter patterns between any of the conditions, $F\ (1, 76) = 1.42$, $p = .24$.

Instructing participants to perform a disconfirming search was not more effective than instructing participants to think of as many rules as possible: The differing instructions did not affect identifying the possible 23 patterns in the rules over the 14 trials, $F(1, 76) = .59$, $p = .44$ (Disconfirm, $M = 15.95$ (69.3%), $SD = 2.88$; Many Rules, $M = 15.39$ (66.9%), $SD = 3.61$). Nor did the different instructions affect eliminating distracter patterns over the 14 trials, $F(1, 76) = .01$, $p = .92$ (Disconfirm, $M = 6.96$, $SD = 3.13$; Many Rules, $M = 7.04$, $SD = 3.23$).

There were no significant interactions between the visual (color or text) and non-visual (instructions) conditions for either identifying patterns in the rule (hits), $F(1, 76) = 1.83$, $p = .18$, or for eliminating distracter patterns (false positives), $F(1, 76) = .06$, $p = .81$.

Overall there was a relationship between identifying patterns in the rule and eliminating distracter patterns. An increase in identifying patterns related to the rule correlates with eliminating distracter patterns, τ_b = -.27 *p* = .001.

Time Efficiency

Three participants' time totals were outliers of more than two standard deviations above the mean and therefore excluded from time analyses (i.e., excluded total time values that exceeded 1700 seconds; 1700 seconds is more than two standard deviations above the mean of 809 seconds).

There was no difference between conditions in time required to determine a rule, $F(3, 73) = .29$, $p = .84$ (see Table 4).

Selection Sequence

As previously described, the triples were initially not visible; however, six boxes were identified by text or color as containing triples that were conforming, partially conforming or nonconforming to a rule. Participants could select up to six triples in any order. The top three selection orders accounted for 67.8% of all searches: 1) Selection of items from top to bottom in the column (List Order – 33.6% of total trials); 2) selection of conforming (CC), then partially conforming (PP), and finally nonconforming sequences (NN); (CCPPNN – 24% of all trials); and, 3) just the conforming triples (CC – 10.2% of all trials). For trial 10 the list order was CCPPNN – which matched both of the top two selection orders; therefore, data from trial 10 were excluded from the selection sequence analyses.

Participants generally selected a search strategy and continued to use the same selection sequence for all trials; as evidenced by the negative correlation between

Table 4. Condition Time Comparisons

Condition	Mean (seconds)	Std. Deviation	N
Disconfirm and Text – I1V1	792.37	257.25	20
Disconfirm and Color – I1V2	836.03	293.14	18
Many Rules and Text – I2V1	842.15	242.48	19
Many Rules and Color – I2V2	770.89	329.72	20
Total	809.28	278.88	77

selecting the two most popular sequences - list order and CCPPNN, τ_b (Kendall's tau_b) = -.59, $p < .001$. However, a multi-attribute analysis of variance (MANOVA) of the top three search sequences with instructions and text color as fixed factors, suggests that only instructions had an effect on search sequences; the CCPPNN selection sequence was used more often with the instruction to disconfirm the rule than the many rules instruction, F (1, 76) = 2.28, $p = .037$ (Disconfirm, $M = 4.15$; Many Rules, $M = 2.23$). However, there was no correlation between the CCPPNN selection sequence and identifying patterns in the rule (hits) $\tau_b = -.05$ $p = .57$ and eliminating distracter patterns (false positives) $\tau_b = -.10$ $p = .24$. There were no significant differences or relationships between the other conditions – list order selection sequence or the CC selection sequence and instructions or color and text conditions.

Search Sufficiency

Given that selection of just one conforming sequence was necessary and sufficient for identifying all patterns in the rule, all but 3 (99.7%) of the 1120 total searches were sufficient for identifying all patterns in the rule. Selection of both conforming sequences and both nonconforming sequences ensured a sufficient search to identify distracter patterns and thereby eliminate them from the rule. Selection of the partially conforming triples was not necessary for a sufficient search. Participants' searches were overwhelmingly sufficient for eliminating distracter patterns (895 of the 1120 total searches, 80%); (see Table 5 and Table 6). There was no significant correlation between any experimental conditions and search sufficiency, $\tau_b = -.02$ $p = .80$.

Search sufficiency also does not correlate with eliminating distracter patterns from the rule, $\tau_b = -.10$ $p = .24$. However, an insufficient search where only the two conforming triples (CC) were selected is related to a failure to eliminate distracter patterns (false positives) $\tau_b = .21$ $p = .01$.

Discussion

The data visualization manipulation of this experiment addressed the question, can increased salience of data characteristics reduce conformation bias and thereby improve performance, with the hypothesis that color can be more effective than text in identifying characteristics of evidence, and therefore replace text for increased search efficiency, sufficiency and decision accuracy. Contrary to the hypothesis, the data do not indicate any significant differences between any color and text conditions. Using color to differentiate characteristics of evidence did not increase search efficiency measured by time. Nor did color affect the order that triples were selected. Both color and text conditions produced equally mediocre results for identifying patterns in the rule; the average for the 14 trials was approximately 66% - 68% over all conditions

The instruction manipulation of this experiment addressed the question, can instructions to structure appropriate hypotheses and sufficiently consider data, reduce confirmation bias and thereby improve performance, with the corresponding hypothesis that instructions to perform a disconfirming search are more effective than instructions to think of as many rules as possible. The results show that instructing participants to use

Table 5. Sufficient Searches by Trial and Search Type

Trial	< 6 selected	All 6 selected	List Order	CCPPNN	Total
1	0	17	31	14	62
2	2	17	28	20	67
3	2	16	28	18	64
4	2	13	31	20	66
5	3	12	28	25	68
6	4	12	27	16	59
7	4	13	25	22	64
8	5	22	22	20	69
9	2	15	24	19	60
10	1	7	55	Same as list order	63
11	4	14	29	17	64
12	3	12	25	25	65
13	3	18	24	19	64
15	3	12	25	20	60
Totals	38	200	402	81	895

Table 6. Insufficient Searches by Trial and Search Type

Trial	CC only	Cs + Ps no Ns	Cs + 1 N	No Cs	Total
1	7	3	7	1	18
2	3	5	4	1	13
3	6	6	3	1	16
4	6	4	4	0	14
5	5	6	1	0	12
6	10	9	2	0	21
7	10	5	1	0	16
8	8	3	0	0	11
9	15	3	2	0	20
10	12	3	2	0	17
11	10	5	1	0	16
12	11	4	0	0	15
13	11	4	1	0	16
15	12	7	0	1	20
Totals	126	67	28	4	225

disconfirming evidence to test the rule was no better or worse than instructing participants to think of as many rules as possible. Neither of the instructions were enough to ensure generation of appropriate hypotheses and sufficient consideration of the data; as evidenced by identifying on average less than 16 of the 23 possible patterns in the rule (69.6%) and failing to eliminate an average of 7 distracter patterns over 14 trials. In sum, the results show no difference in search sufficiency, accuracy or efficiency between the instruction conditions. The only difference found between the instruction manipulations was the favoring of the CCPPNN selection sequence when participants were instructed to consider disconfirming data. Although, the instructions may have affected the order evidence was considered, the CCPPNN selection sequence does not correlate with any change in identifying patterns in the rule or eliminating distracter patterns.

Perceived Value

The instructions to test the rule were intended to encourage participants to consider disconfirming evidence – perhaps by facilitating a switch from associative processing to rule-based reasoning. A closer look at participants' descriptions of their strategies revealed that only 25 (31.3%) of the 80 participants mentioned the value of disconfirming evidence for testing their rules. Of those 25 participants, 12 (48%) were in the condition with instructions to test the rule using disconfirming evidence. An almost equal number – 13 (52%) – of participants were in the many rules condition. With no real difference between conditions, clearly the instruction to use disconfirming evidence did not increase the value participants placed on disconfirming evidence. In fact, the data possibly indicate the opposite effect: 7 of the participants explicitly stated there was no value in disconfirming evidence, with statements such as, "…as I went along, I realized I only needed to uncover the conforming numbers." All 7 participants who stated there was no value in disconfirming evidence were in the condition instructed to test the rule using disconfirming evidence.

The remaining 55 (68.8%) participants who did not mention disconfirming evidence described their strategies in terms of finding the patterns for the rule with no mention of the data characteristics (conforming or nonconforming) that they considered. Participants overall disregard for disconfirming evidence is reflected in the lack of correlation between search sufficiency and accuracy. Even though participant's searches were overwhelmingly sufficient to reveal all the evidence needed to identify patterns and eliminate distracter patterns, it seems most participants lacked an understanding of how to use the evidence; as reflected in the failure to eliminate an average of seven distracter patterns from the rules over 14 trials.

This experiment did not restrict how many triples could be selected. While the unrestricted search resulted in selections sufficient to discover the necessary evidence it is difficult to determine what data were actually considered – even after reviewing participant strategy descriptions. Limiting the ability to select the data and asking participants to state what the rule could be after every selection might facilitate judging the perceived value of the data and what data were actually considered. In addition, other methods of increasing the salience of data patterns might support discovery of the rule and aid eliminating distracter patterns. Experiment 2 explored these possibilities.

EXPERIMENT 2 – SALIENCE AND PERCEIVED VALUE OF EVIDENCE

Introduction

Salience

Previous studies using the Wason 2-4-6 task demonstrated that in addition to feedback on the triples' adherence to the rule, the discovery of the rule was facilitated by displaying a graphical representation of the participant's triple (Vallée-Tourangeau & Payton, 2008). Experiment 2 continued to explore increasing salience of data characteristics to reduce confirmation bias and thereby improve performance. Specifically by using a graph to investigate the question, can increasing the salience of data patterns reduce confirmation bias and increase search efficiency, sufficiency and decision accuracy?

This experiment included the presentation of a graph that reflected the three patterns associated with the rules: Providing line charts to increase the salience of ascending and descending patterns, the differences between numbers, and the patterns of odd or even numbers. The hypothesis was that increasing the salience of patterns related to the user's task by presenting a graph reflective of these patterns reduces confirmation bias and the cognitive effort for comparisons; resulting in, a) increased search sufficiency – measured by selection; b) increased decision accuracy; and, c) increased efficiency – measured by time. The hypothesis was related to the premise that increased search sufficiency and increased accuracy may result from a switch to more rule-based processing to generate hypotheses and sufficient consideration of the data; higher efficiency may result from supporting associative processing.

Perceived Value of Evidence

The results of Experiment 1 showed a lack of correlation between search sufficiency and accuracy; this raised questions regarding the perceived value of evidence. Pirolli (2007) explored search patterns and perceived value of data. Pirolli's information foraging theory compares the search for information to an animal foraging for food. The analogy posits that the forager will pick off the most profitable items first and will limit their investment where they perceive little gain. Consistent with this analogy, the search selection sequences reveal the evidence the searcher considers the most profitable. The items selected first are perceived as having the most value. Items not selected are perceived to have less value or provide little to no gain.

This experiment further explored the perceived value and consideration of disconfirming data. Scarcity was manipulated by limiting the ability to select evidence, with the objective of encouraging a more profitable selection sequence. To facilitate interpreting the perceived value and consideration of the data, participants were asked to state what the rule could be after every selection; in addition, they were asked about their strategies after every trial and at the end of the 15 trials.

Method

Participants

Participants were recruited from the University of Utah psychology undergraduate student participant pool. Data were analyzed from 42 participants ranging in age from 18 to 40 years ($M = 23$). There were 24 (57%) female participants and 18 (43%) male.

Exclusions

The procedure for testing color vision and determining data exclusion was identical to Experiment 1. Participants' data were excluded if the individual could not distinguish between the colors used later in the experiment to identify data characteristics.

In Experiment 1 participants selected all the triples and then entered the rule once for every trial. For this experiment the participants were asked to enter what they thought the rule could be after selecting each triple. There was only one rule per trial; however, entering a rule after every triple (up to 4 triples selected) was confusing to some participants. Ten potential participants persistently misinterpreted the task as defining a new, separate rule for each triple. Their data were excluded from analyses. (See the Data Coding section for a more complete explanation of how the decision to exclude was determined.)

Design

The between-subjects design examined the impact of the independent variable with the two levels of salience: Presence of graph and absence of graph. The two experimental conditions were used to examine the question, can increasing the salience of data patterns reduce confirmation bias and increase search efficiency, sufficiency and decision accuracy?

The software was modified for the conditions of this experiment. As in Experiment 1, six boxes containing triples were presented in a column – triples were initially not visible. When selected by the participant a triple was revealed in that box. The results of Experiment 1 showed color to be no more or less effective than text in identifying triples. Colors were used instead of text in this experiment to consistently correlate identifying the triples in the boxes with the same triples charted in the graph. Each of the boxes were identified as “conforming,” “nonconforming,” or “partially conforming” to a rule using color coding – green, yellow and red – to match rule conformation (i.e., green = conforming; yellow = partially conforming; red = nonconforming); (see Figure 8).

Identical to Experiment 1, the boxes containing triples were positioned in pseudo-random order, with a different order for every trial but consistent between conditions for each trial (i.e., the positions in the list for conforming and nonconforming triples differed between trials 1 through 15; however, the order of conforming and nonconforming triples in trials 1 through 15 were the same for all participants); (see Table 2).

In the first condition (no graph) when each colored box was selected the associated triple was revealed (see Figure 8). In the second condition (graph), when each colored box was selected, in addition to revealing the triple, the triple was charted as a line on a graph to the right of the boxes (see Figure 9). The colors of the line chart corresponded to the colors identifying the conformance of the triple with the rule (green, yellow or red). In addition, even numbers were marked on the graph in blue and odd

Figure 8. Example trial, nongraph condition

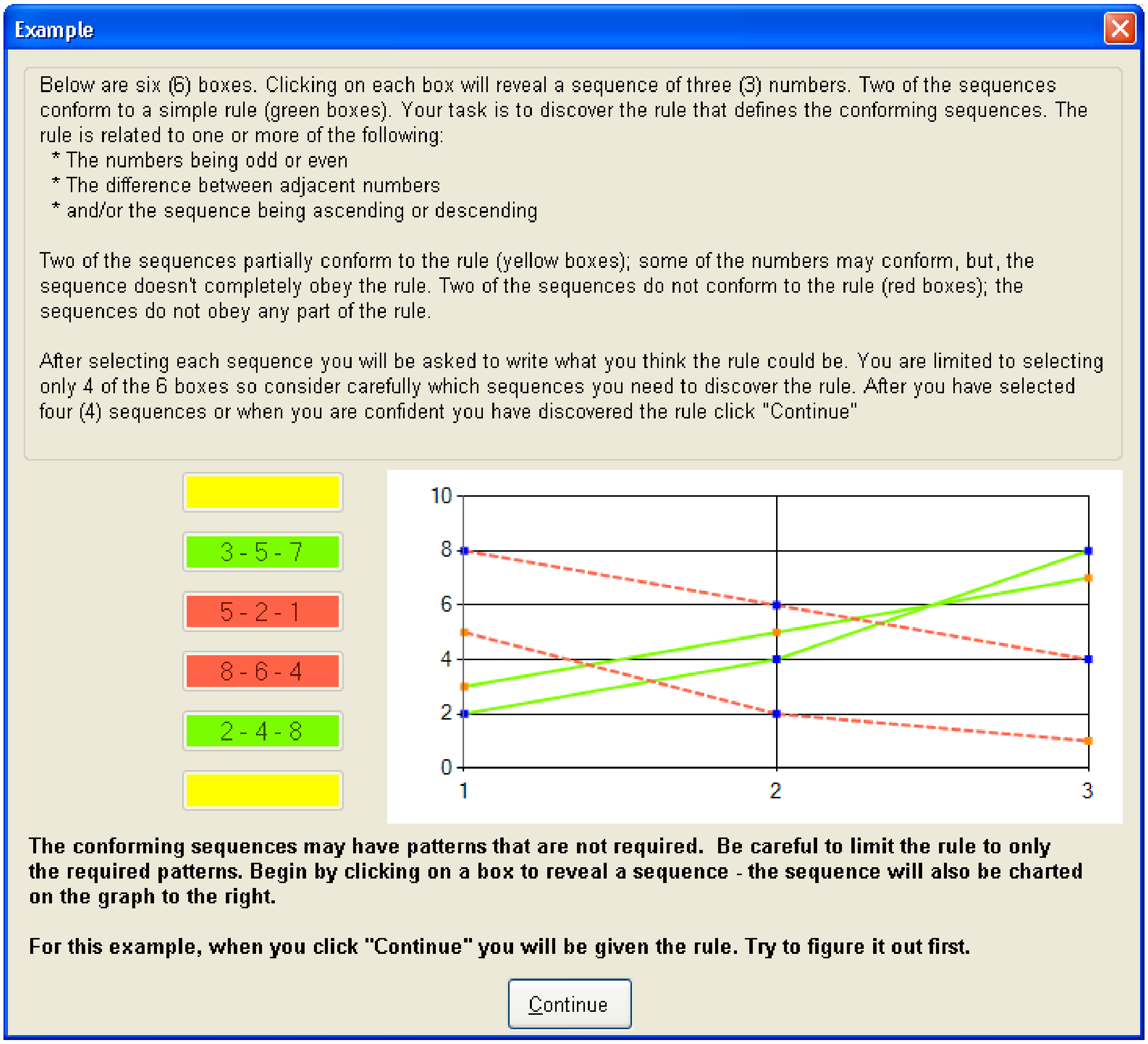

Figure 9. Example trial, graph condition

numbers were marked in orange (see Figure 10). Graphing was used to increase the salience of ascending and descending patterns, differences between numbers and odd/even patterns.

Procedure

As in Experiment 1, participants were greeted and given a color vision screening test. A research assistant (RA) logged in and entered a predetermined unique participant ID. The RA directed the participant to read the instructions aloud. Participants were encouraged to ask questions while they completed an example trial. Several participants questioned whether each triple was associated with a separate rule. The RAs responded that each trial had only one rule and then re-read the task instructions to the participant, emphasizing that the participant's task was "to discover the rule that defines the conforming sequences."

Participants interacted with the application by clicking on up to four of the six initially empty boxes. The boxes were identified by color as containing triples that, 1) conformed to the rule, 2) did not conform to the rule, or 3) partially conformed to the rule (two triples of each type). Participants could select the boxes in any order

After selecting a box, the triple appeared in the box and a dialog box appeared with instructions to enter what they thought the rule could be (see Figure 11). After

Figure 10. Graph explanation

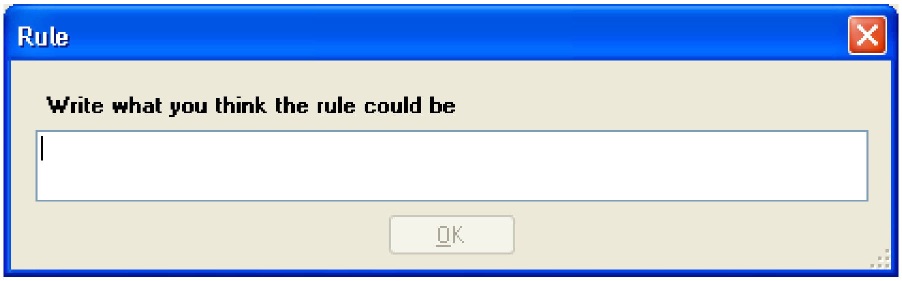

Figure 11. Prompt for entering the rule after the first selection

typing in a rule, participants clicked on the "OK" button and continued selecting boxes to reveal up to four triples.

Selecting the "Continue" button displayed the text, "Write the strategy you used to determine the rule. Click 'Continue' when you are finished." Participants then entered the strategy they used to determine the rule. Selecting the "Continue" button again caused the correct rule to appear (see Figure 2). After dismissing the dialog box containing the correct rule, the screen was reset to the next trial. After 15 trials the participants were presented with a set of questions requiring input into three response fields (see Figure 12 and Figure 13):

- The first response field required selecting a "yes," "no" or "unsure" to answer the question, "Did your strategy change as you did more tasks?"
- The second response field required entering text to answer the questions of why their strategy changed or did not change and if their strategy changed, how it changed.
- The third response depended on the experimental condition. The participants responded to "Did the colors and/or graph help? Why or why not?" (graph condition); or, "Did the colors help you? Why or why not?" (nongraph condition).

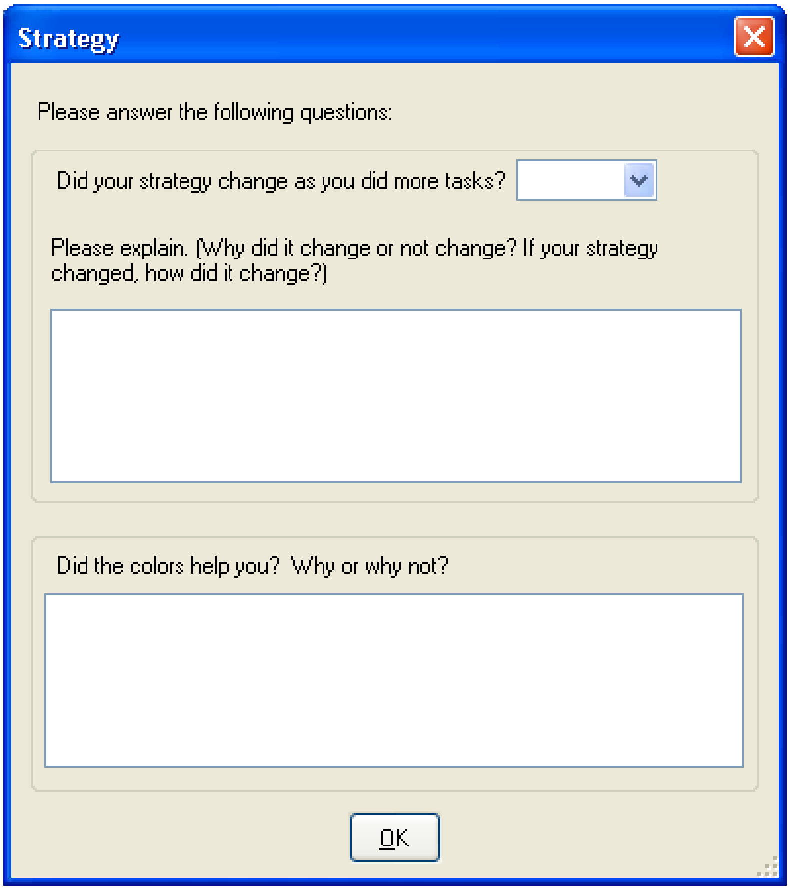

Figure 12. Strategy questions for nongraph condition

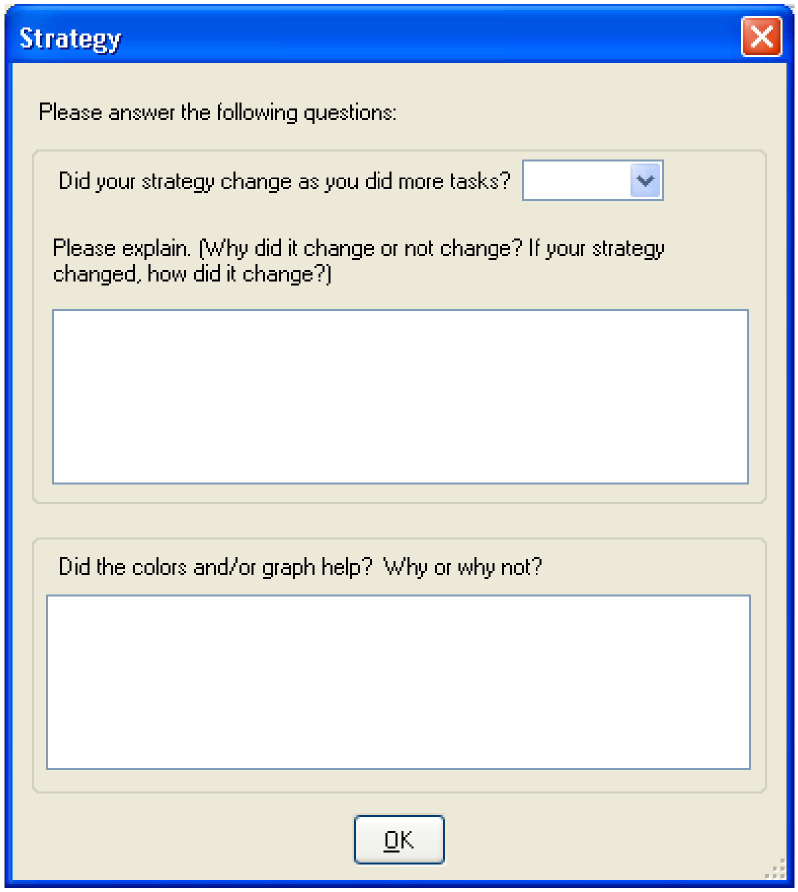

Figure 13. Strategy questions for graph condition

After responding to the questions and selecting "OK," the participants answered demographic questions – identical to the questions in Experiment 1 (see Figure 3).

Measurements

The following measurements were recorded in a database for each trial:

Time. The time was recorded for participant actions:

- When each triple was selected (made visible)
- When the rule was entered for each triple
- When the participant selected "Continue" after entering their strategy for each trial

These measurements allowed comparison of the time for trials between and within participants, and for each experimental condition. Search efficiency was measured by time.

Rule. Participants were instructed to enter what they thought the rule could be after selecting each triple. The triples and rules were identical to the triples and rules used in Experiment 1.

Search sequence. Identical to Experiment 1, when each triple was selected and made visible, the type of the triple – conforming, nonconforming or partially conforming – was recorded as well as the time of selection. This facilitated determining the search sequence.

Strategy. The participant entered a description of the strategy they used to determine each rule. The example trial differed from Experiment 1 in that the description of potential strategies was the same for both conditions of the experiment: "There are multiple strategies for determining the rule; one strategy is to select the first conforming sequence (3-5-7) and think of possible rules for that sequence. You could hypothesize the rule is odd numbers with adjacent numbers differing by 2 and ascending numbers. Looking at other sequences can confirm or disconfirm that hypothesis. Other strategies may be more or less effective. You are encouraged to develop your own strategies."

Data Coding

The results of Experiment 2 were coded using the same methods and criteria as Experiment 1 with exceptions and details as described below.

Rules

The process for coding the "hits" and "false positives" in the participant rule responses was performed as described in the General Data Coding section. However, because the participants entered a rule after selecting each triple instead of coding only one rule per trial four attempts to define the rule were coded for each trial. The selection sequences were also coded as described in the General Data Coding section.

Times

All times were calculated electronically using values captured and stored by the software application. Time to determine the rule was measured from when the participant selected the first triple to when the participant selected the "Continue" button, before they entered their strategy for that trial. Therefore, time measurements included the time to write the rule.

Data Exclusion

Giving participants four attempts to define the rule confused some participants. Some participants interpreted each triple as having a separate rule. During the example trial when participants questioned whether each triple was associated with a separate rule the RAs responded that each trial had only one rule and then re-read the task instructions to the participant; emphasizing that the participant's task was "to discover the rule that defines the conforming sequences."

Participants demonstrated an understanding of the task by entering what they thought the rule was on the first attempt and on subsequent attempts repeated the same rule or added or dropped a pattern. For example, a participant, with the ID = 456, on trial 11 first selected the conforming triple "7-5-3" and entered the rule "descending odd numbers." After selecting the second conforming triple "9-5-3" the participant re-entered the same rule – "descending odd numbers." After selecting the first nonconforming triple "2-5-8" the participant correctly changed the rule to simply "descending."

Ten potential participants (19%) failed to understand that there was only one rule per trial. Their responses defined a separate rule for each triple selected regardless of whether the triple was identified as conforming or nonconforming. For example, on trial 11 a potential participant, with the ID = 457, first selected the conforming triple "9-5-3" and entered the rule "descending, odd." After selecting the other conforming triple "7-5-3," (s)he reentered the same rule "descending, odd." However, after selecting the first nonconforming triple "2-5-8" (s)he changed the rule to "ascending." After selecting the second nonconforming triple "1-3-5" (s)he changed the rule to "ascending, odd." By defining a rule based on the patterns in the nonconforming triples the participant demonstrated a misunderstanding of the task. All data from potential participants who demonstrated this lack of understanding were excluded from analyses.

Results

Search Sufficiency

Selecting only one conforming triple was sufficient to identify all the patterns related to the rule. Participants had enough evidence to identify all the patterns in the rule for 99.8% of the trials – 1 participant selected only a partially conforming triple in one trial.

Both conforming triples were selected in 619 (98%) of the 630 total trials. However, selecting all of the conforming and nonconforming triples may be necessary and was always sufficient for eliminating distracter patterns. A sufficient search for eliminating distracter patterns occurred in only 68 (11%) of the 630 trials. There was no relationship between graph and no-graph conditions and search sufficiency τ_b =-.03. p = 87. Selection of the partially conforming triples was not necessary for a sufficient search.

Accuracy

A one-way ANOVA with graph, no graph as the between subjects factor indicated that graphing the triples had no significant effect on identifying the possible 24 patterns in the 15 rules (on the last attempt to define the rule for each trial), $F(1, 40) = .77$, p =.39 (overall M = 17.27 (71.8%), SD = 2.90; no graph, M = 16.88 (70.3%), SD = 3.22; graph, M = 17.67 (73.6%), SD = 2.56). There was also no significant difference in eliminating distracter patterns between conditions, $F(1, 40) = .80$ $p = .38$, (overall, M = 12.52, SD = 2.76; no graph, M = 12.14, SD = 2.69; with graph, M = 12.90, SD = 2.84).

Ascending and descending patterns were perhaps the most salient patterns represented on the graph. Not all rules included ascending or descending values as a target pattern. In the trials where ascending or descending was a target pattern a post-hoc t-test indicated no difference between the graph and no graph conditions in identifying the possible 13 patterns in seven rules, t (40) = -1.08, p = .29, (no graph, M = 9.62 (74%), SD = 1.86; graph, M = 10.24 (78.8%), SD = 1.87). In the eight trials where ascending or descending was a distracter pattern, the graph condition reflected significantly more difficultly in eliminating distracter patterns from the rule (on the last attempt), t (40) = -2.49, p = .02, (no graph, M = 4.81, SD = .81; graph, M = 5.67, SD = 1.35).

Overall, participants identified on average a maximum of 13.95 distracter patterns, SD = 3.20, and failed to eliminate 12.52 of those distracter patterns, SD = 2.76 on the last attempt over the 15 trials. Although the reduction was numerically small, a paired samples t-test showed the reduction was statistically significant, t (41) = 8.20, p <.001. Based on the last attempt per trial, participants overall accuracy for identifying patterns in the rule while eliminating all the distracter patterns was 3.13 of the 15 trials (20.9%).

Time Efficiency

For both graph and no graph conditions, participants' average time per trial for the first three trials was significantly greater than the average time per trial of the following 12 trials, t(41) = 7.6, p <.001 (first 3 trials, M = 148.73 s, SD = 61.54; last 12 trials, M = 82.45 s, SD = 35.12); (see Figure 14).

However, participants in the graph condition required significantly more time on the first three trials than participants with no graph, t(40) = -2.11, p = .04, (no graph, M = 388.44 s, SD = 159.57; graph, M = 503.95 s, SD = 193.44). There was not a significant difference in time between the graph and no graph conditions for the remaining 12 trials,

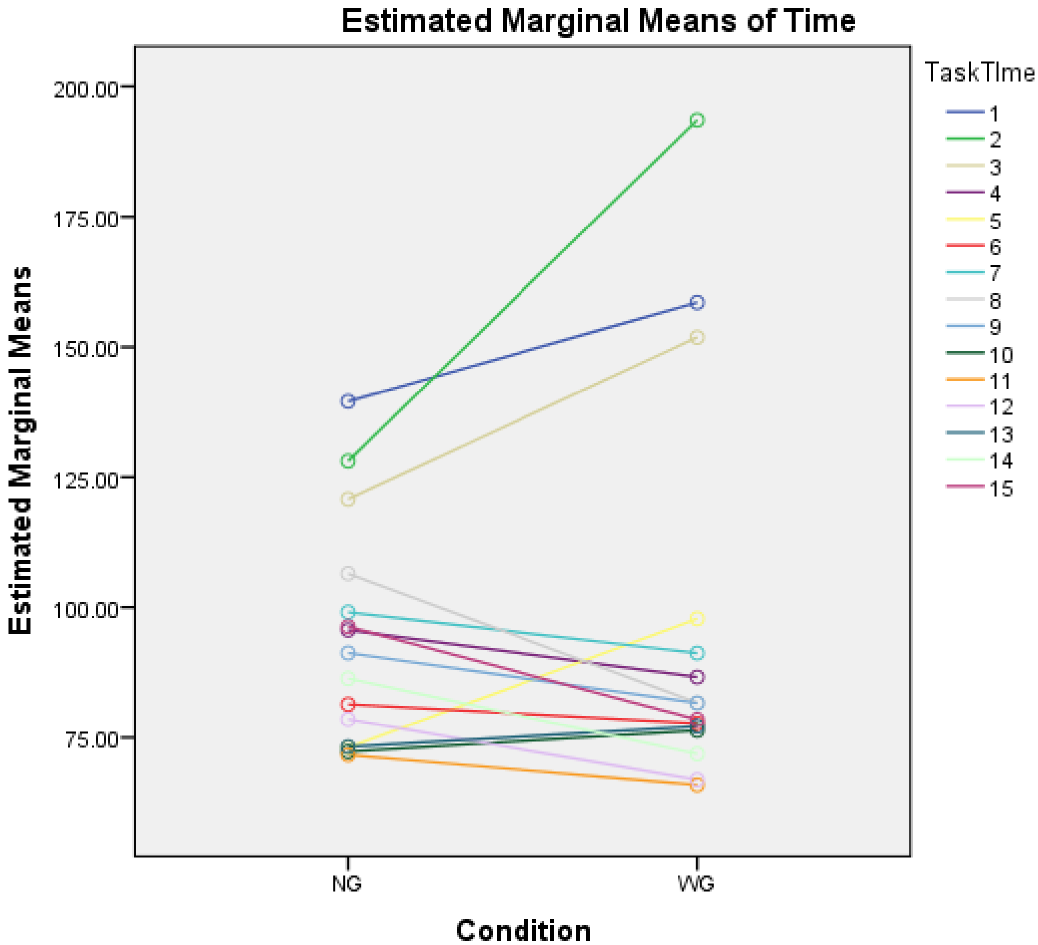

Figure 14. Time differences between graph conditions

$t(40) = -.26$, $p = .79$, (no graph, $M = 1025.25$ s, $SD = 532.13$; graph, $M = 953.48$ s, $SD = 279.61$).

Perceived Value of the Graph

Participants in the graph condition were asked at the end of the session if they found the colors and/or graph helpful; participants in the no graph condition were only asked if they found the colors helpful (see Figure 12 and Figure 13). Of the 21 participants in the graph condition 7 (33%) found the graph helpful and 8 (38%) did not find it helpful or did not use the graph – 6 (29%) participants did not comment on the usefulness of the graph. Participants' responses on the usefulness of the graph included, "The visual of the graph allowed me to judge quicker and see correlations between the numbers and their sequences rather than having to envision them or do mental math" and "The graph only helped a little but I used it more as I did more tasks."

There was little numerical difference in identifying the possible 24 patterns in the 15 rules between participants who perceived the graph as helpful and those who did not (on the last attempt); (graph helpful, $M = 17.14$ (71.4%), $SD = 2.36$; graph not helpful, $M = 18.12$ (75.5%), $SD = 3.10$). Overall there was no relationship between the perceiving the graph as helpful and detecting patterns related to the rule, $\tau_b = -.09$. $p = .23$. There was also little numerical difference in eliminating distracter patterns between participants who perceived the graph as helpful and those who did not (on the last attempt); (graph helpful, $M = 13.86$, $SD = 4.10$; graph not helpful, $M = 13.00$, $SD = 1.51$). In addition perceiving the graph as helpful had no correlation to eliminating distracter patterns, $\tau_b = -.06$. $p = .32$.

Perceived Value of Evidence

As previously described, the triples were initially not visible; however, six boxes were identified by color as containing triples that were conforming, partially conforming or nonconforming to a rule. Participants could select up to four boxes in any order. The top four selection sequences were, 1) both conforming triples then both partially conforming triples (CCPP – 109 trials or 17.3% of all trials); 2) both conforming triples then one nonconforming and one partially conforming triple (CCNP – 95 trials or 15.1% of all trials); 3) both conforming triples (CC – 92 trials or 14.6% of all trials); and, 4) both conforming triples then one partially conforming triple and one nonconforming triple (CCPN – 74 trials or 11.7% of all trials). The top four selection sequences accounted for 370 trials or 58.7% of all trials (see Table 7).

All of the top four selection sequences included selecting both of the conforming triples first. Two other selection sequences included selecting both of the conforming triples first (i.e., CCNN and CCN), for a total of 450 of the 630 trials (71%). Participants selected triples in list order for 13 trials or 0.019% of all trials.

Considering selection of triples without regard to order, the following were the most common selections:

- Both conforming triples, one nonconforming triple and one partially conforming triple (CCPN, CCNP, CNCP, CNPC) – 184 trials (29%)
- All conforming and partially conforming triples (CCPP, PPCC) – 132 trials (21%)
- All conforming and all nonconforming triples (CCNN, NCCN, CNNC, CNCN) – 68 trials (11%)

Table 7. Experiment 2 – Selection Sequences

Selection Sequence	Participants Using	Min	Max	Mean	Std. Deviation	Total trials	% of Trials
CC	18	1.00	13.00	5.11	4.56	92.00	14.6%
CCNN	13	1.00	12.00	4.00	3.67	52.00	8.3%
CCPP	22	1.00	15.00	4.95	4.82	109.00	17.3%
CCNP	22	1.00	10.00	4.32	3.12	95.00	15.1%
CCPN	25	1.00	12.00	2.96	2.68	74.00	11.7%
CNCP	10	1.00	9.00	2.10	2.51	1.00	3.3%
CCN	9	1.00	6.00	3.11	2.09	28.00	4.4%
CNCN	6	1.00	2.00	1.67	.52	10.00	1.6%
CNNC	2	1.00	3.00	2.00	1.41	4.00	0.6%
CNPC	5	1.00	10.00	2.80	4.02	14.00	2.2%
PPCC	3	1.00	15.00	7.67	7.024	23.00	3.7%
NNCC	0	0	0	0	0	0	0
NCNC	0	0	0	0	0	0	0
NCCN	2	1.00	1.00	1.00		2.00	0
List Order	3			1.00		3.00	0.004
List Order Trial10	10			1.00		10.00	0.015%
Totals	150					537 of 630	85.24%

Discussion

This experiment explored the impact of salience and perceived value and consideration of disconfirming evidence. The findings related to salience are discussed first and then the findings regarding perceived value.

Salience

Increasing the salience of data patterns by charting on a graph was expected to increase search sufficiency, decision accuracy and efficiency. The findings are discussed in turn for each of these measures.

Search Sufficiency and Decision Accuracy

The graph condition failed to facilitate an increase in search sufficiency. Overall, only 11% of the searches included selecting all of the disconfirming evidence. Since all the necessary data were rarely selected it is not surprising participants failed to eliminate an average of 12.5 distracter patterns over the 15 trials. Accuracy where all distracter patterns were eliminated averaged only 20%. Considering only trials with sufficient selection, the accuracy when all distracter patterns were eliminated was not significantly different (p = .38) at 19% over all trials. Therefore, even when the participants selected sufficient evidence they rarely eliminated all distracter patterns.

Despite the purported ability of graphical representations to increase comprehension of trends and relationships (see Few, 2006; Tufte, 2000; Wainer, 2005), the present results indicate that the proper interpretation of a graph's values cannot be assumed. Instead of increased decision accuracy there was no difference in identifying patterns related to the rule in the graph versus no graph conditions.

Efficiency

Charting values on a graph not only failed to increase efficiency, the graph condition results show decreased efficiency. The increased time to consider the graph was reflected in the average time for the first three trials; moreover, the increased investment of time was not offset by any gains in efficiency in the subsequent trials. The hypothesis that increasing the salience of patterns related to the user's task reduces cognitive effort for detecting and comparing patterns was not supported.

It is unlikely that the graphs in this experiment promoted switching from associative to rule-based processing. In fact, the increased salience of ascending or descending patterns in the graph may have strengthened associative processing in the graph condition, as evidenced by the increased failure to eliminate ascending or descending distracter patterns from the rule.

The presence of the graph did not support the generation of hypotheses or consideration of disconfirming data. In sum, the results of this experiment provide neither evidence that increasing the salience of data patterns reduces confirmation bias nor reduces cognitive effort for comparisons.

The lack of benefit from the graph could be due to poor representation in the graph and/or participants' poor problem-solving skills and misunderstanding of task requirements (Jarvenpaa, Dickson, & DeSanctis, 1985). Participants' lack of understanding the value of disconfirming evidence might also have contributed to the failure to use the graph to improve accuracy and efficiency.

Perceived Value of Evidence

Based on information foraging theory (Pirolli, 2007) one could conclude that conforming data were considered the most valuable since conforming triples were selected first in an overwhelming number of searches (71%). In addition, disregarding order, all conforming triples were selected in 98% of the trials. In contrast, both nonconforming triples were selected in only 68 trials (11%). Interestingly, ambiguous evidence was considered more valuable than potentially disconfirming evidence – based on the frequency of selecting both partially conforming triples (132 trials – 21%) nearly twice as often as selecting both nonconforming triples (68 trials – 11%).

These findings are consistent with the propensity resulting from biases for insufficient consideration of the data. This experiment probed for insight into perceived value and consideration of data by asking participants about their strategies.

Strategy

Participants were asked after every trial for the strategy they used to determine the rule. After the 15 trials participants were asked if their strategy changed and if it changed, how did it change? Participant responses indicated a varied understanding of disconfirming evidence; ranging from complete dismissal of nonconforming triples to a stated recognition of their value. Fourteen of the 42 participants (33%) stated they disregarded the nonconforming – red – sequences or did not find them helpful. Participants explained their disregard of nonconforming triples as typified by: "The green always told you the rules so I mainly kept checking those instead of wasting time and checking the red that wouldn't tell you anything." "I also decided right off the bat to not even use the red boxes, and went for the yellow first and then the green to finalize the rule." "I knew in order to get the strategy a completely wrong answer wouldn't help so I never picked the red ones."

Ten of the 42 participants (24%) stated the nonconforming triples were helpful. However, their understanding was sometimes incomplete: "At first, I went green, green, yellow, red. As I went, I found that green, then red was helpful, because you can quickly disconfirm many things via the discrepancy between them… Using 2 red boxes I found to be useless; you get all the information you would need from one red box, and then you don't get to choose a more helpful green or yellow." A few participants seemed to arrive at what could be a successful strategy, typified by, "After I did a few trials, I found that it was easier to use the red boxes to eliminate guesses, than to use the yellow boxes which may or may not conform to all the rules, and ultimately ended up confusing me more."

Search Sufficiency

None of the participants conducted a sufficient search for all 15 trials. Twenty-five participants (59.6%) had zero sufficient searches. Of the 10 participants who stated they recognized value in the nonconforming triples, totally sufficient search patterns of selecting two conforming and two nonconforming triples were followed in only 38 of their 150 trials (25%); (CCNN, CNCN, CNNC, NCCN). They were more likely to select both conforming triples, one partially conforming and one nonconforming triple (CCNP, CCPN, CNPC, CNCP) for a total of 64 of 150 trials (43%). Therefore, even though they stated a positive perceived value for disconfirming evidence, their searches were nevertheless deficient. This search deficiency is reflected in the lack of correlation between finding the nonconforming triples helpful and eliminating distracter patterns τ_b =

.11. $p = .20$. Therefore, perception of value is not reliable; even when disconfirming evidence was perceived as helpful there was no reduction in errors.

This lack of correlation between perceived value and selection prompted a closer look at individual's searches and stated strategies. The highest level of search sufficiency was 12 of the 15 trials – achieved by only 1 participant. However, that participant identified all the patterns in the rule and eliminated all the distracter patterns in only one trial. Even though the participant discovered all 24 patterns in the 15 rules, (s)he failed to eliminate 14 distracter patterns. The participant explained the strategy as "… choosing the green then red boxes. The one time I changed it up, it left me confused (choosing a yellow box.) So I stuck with what worked." Clearly their perception of "what worked" did not include eliminating distracter patterns. The participant with the second greatest number of sufficient searches (11 of 15) performed slightly better by eliminating all distracter patterns in 4 of the 15 trials.

The participant with the best overall results did not perform a single sufficient search. Their search pattern of choice was a combination of two conforming triples, then one nonconforming triple and finally one partially conforming triple (CCNP). This selection pattern started on trial 1 and was followed for all 15 trials. The participant described the strategy as "…looked at the green box, determined what the rule could be, confirmed with second green box, compared it with red box, and looked for similarities with hypothesis with the yellow box to confirm once more." This strategy allowed them to identify 18 of the 24 patterns (75%) in the 15 rules; however, they still failed to eliminate six distracter patterns over the 15 trials.

In summary, there was a lack of correlation between finding the nonconforming triples helpful and eliminating distracter patterns. Examining participants' strategies and search patterns revealed that a sufficient search strategy did not ensure eliminating distracter patterns, and the best overall results were obtained without a sufficient search. However, the suboptimal values of the "best results" clearly reflect the insufficient search.

Disconfirming Evidence and Decision Accuracy

Within Experiment 2, there was a lack of correlations between perceiving nonconforming data as valuable, search sufficiency and eliminating distracter patterns. This raised the question, does selecting disconfirming evidence make a difference in decision accuracy? This question led to a comparison of Experiments 1 and 2. In Experiment 1 participants could select up to all of the triples. Participants' searches in Experiment 1 were overwhelmingly sufficient for eliminating distracter patterns (80%). Experiment 2 limited the search selection to four triples. Limiting the number of selections had a dramatic effect on selection sequences. To illustrate this effect, the selection of all items in the list from top to bottom (list order) was one of the top selection patterns in Experiment 1 (31.2% of all trials); the list order selection pattern all but disappeared in Experiment 2 (0.019%) trials; selecting only the conforming sequences (CC) increased from 10.2% in Experiment 1 to 14.6% of all trials in Experiment 2. Participants' searches in Experiment 2 were overwhelmingly insufficient for eliminating distracter patterns; only 11% of searches were sufficient for eliminating distracter patterns.

Despite there being no correlation between participants declaring that nonconforming triples were helpful and actually eliminating distracter patterns from the rule, there was value in selecting nonconforming triples: While there was no difference between Experiment 1 and Experiment 2 in accurately identifying the patterns in 14 rules, $t(120) = 1.46$, $p = .15$ (Experiment 1, $M = 9.42$ (67.3%), $SD = 2.04$; Experiment 2, $M = 9.96$ (71.1%), $SD = 1.73$); the decrease between Experiments 1 and 2 in searches sufficient to eliminate distracter patterns does correspond to a significant increase in failure to eliminate distracter patterns in 14 rules (on the last attempt), $t\ (120) = -7.22$, $p < .001$ (Experiment 1, $M = 7.00$, $SD = 3.16$; Experiment 2, $M = 11.10$, $SD = 2.58$).

Therefore, regardless of the participants' stated perceived value, selecting disconfirming evidence did result in increased use of the disconfirming data. This finding raised the question, if selection of all the necessary disconfirming evidence cannot be assumed in self-directed searches, can presenting sufficient disconfirming evidence reduce confirmation bias and thereby improve performance?

A partial explanation of the disconnect between the positive perception of the disconfirming evidence and decision accuracy may be that selecting ambiguous evidence – partially conforming triples – was more compelling to most participants than selecting disconfirming evidence. Perhaps because the ambiguous partially conforming evidence was perceived as disconfirming. This suggests the question, can removing irrelevant ambiguous evidence also improve performance?

EXPERIMENT 3 – NECESSARY AND SUFFICIENT EVIDENCE, AND ORDER

Introduction

Necessary and Sufficient Disconfirming Evidence

The results of Experiments 1 and 2 revealed four potential limitations of self-directed searches. 1) The selection of necessary disconfirming evidence cannot be assumed, regardless of the perceived value of disconfirming evidence. 2) The selection of sufficient evidence does not ensure accuracy; however, 3) insufficient selection of disconfirming evidence results in lower accuracy. 4) Selection patterns indicate that ambiguous evidence is considered more valuable than potentially disconfirming evidence. All four limitations reflect tendencies to not select sufficient relevant evidence and to give undue consideration to irrelevant evidence.

In response to these limitations Experiment 3 considered two questions: 1) Can presenting sufficient disconfirming evidence reduce confirmation bias and thereby improve performance; and 2) can removing irrelevant ambiguous evidence also improve performance?

Order

Presenting necessary and sufficient evidence without requiring a search raises other concerns, including order effects. The order that evidence is presented has been shown to affect human cognition: Items presented first – primacy – and last – recency –

are easier to recall than items presented in the middle of a sequential presentation (Ebbinghaus, 1913). These order effects can impact decision making in multiple contexts, including legal and healthcare (Bergus, Chapman, Levy, Ely, & Oppliger, 1998; Walker, Thibaut, & Andreoli, 1972).

In studies of clinical diagnoses some researchers argue for a recency effect (Bergus et al., 1998) – where the evidence presented last disproportionately affects the diagnosis. Other researchers argue there is a primacy effect – the diagnosis is more likely to be based on the first evidence presented. However, the primacy effect may disappear with more experienced participants (Cunnington, Turnbull, Regehr, Marriott, & Norman, 1997); such experience based effect is consistent with Wang and colleagues' (2000) finding that "order effects in belief revision exist at the early stage of training when the confidence level is low and they tend to diminish and disappear later when the confidence increases." In addition, whether the impact of data is greater when presented first or last depends on many factors including the complexity of the stimuli, length of the series of evidence items, and the data's subjective value for an individual (Hogarth & Einhorn, 1992).

The variation in participants, tasks and measurements for prior order effect studies does not provide definitive answers. Indeed, the effect of the temporal presentation of data on hypothesis generation remains an ongoing question (see Englund & Hellström, 2012; Lange, Thomas, & Davelaar, 2012; Rebitschek, Scholz, Bocklisch, Krems, & Jahn, 2012).

In summary, in the previous two experiments participants were allowed to select data in any order. In this experiment the order of data presentation was controlled, allowing direct comparisons between order conditions. This explored the question, how does evidence order affect the generation of hypotheses? Specifically, after viewing initial evidence, does providing evidence that may confirm the initial hypothesis solidify the decision so that disconfirming evidence is not adequately considered at a later time? Conversely, does providing evidence that may disconfirm the hypothesis before offering confirming evidence promote sufficient consideration of the evidence? What effect do other orders of presentation have on the accuracy of the decision? The hypotheses were that presentation order affects hypothesis generation and corresponding decision accuracy in the following ways:

- Providing confirming evidence first may solidify the hypothesis so that disconfirming evidence is not adequately considered – thus reducing decision accuracy (CCNN order).
- Conversely, providing disconfirming evidence before offering confirming evidence may promote sufficient consideration of the evidence; however, providing disconfirming evidence too quickly may hinder the formation of a complete hypothesis (NNCC order)
- Providing some confirming evidence, then disconfirming, and then more confirming evidence may increase the correct identification of patterns and support eliminating patterns that are not part of the rule (CNNC and CNCN). This may promote the most accurate decisions.
- Providing some disconfirming evidence then confirming evidence will not be significantly different than confirming then disconfirming (NCCN, NCNC).

Method

Participants

Participants were recruited from the University of Utah psychology undergraduate student participant pool. Each participant individually completed one experiment in one session. Data were analyzed from 130 participants ranging in age from 18 to 45 years (*M* = 23). There were 77 female (59%) and 53 male (41%) participants.

Exclusions

As with previous experiments, all potential participants were tested for color vision and their data were excluded from analyses if they could not distinguish between the colors used to identify data characteristics.

Consistent with Experiment 2, the participants were asked to enter what they thought the rule could be after selecting each triple. Again this confused some participants. Thirty-one potential participants (19% of participants) misinterpreted the task as defining a new, separate rule for each triple. Their data were excluded from analyses.

Design

The one factorial between-subjects design examined the effect of the independent variable (presentation order). The conditions were the six permutations for the presentation order of the two conforming and two nonconforming triples (i.e., CCNN, CNNC, CNCN, NNCC, NCCN, NCNC). The conditions were used to examine the question, how does evidence order affect the generation of hypotheses?

Consistent with Experiment 2, the triples were identified by color only and the instructions did not vary – all participants were instructed to discover the rule, with only a basic strategy provided. When each triple was presented the triple was also charted as a line on a graph to the right of the boxes. The display of the graph was identical to the graph condition in Experiment 2: The colors of the line chart corresponded to the colors identifying the conformance of the triple with the rule (green, yellow or red). Even numbers were marked on the graph in blue and odd numbers were marked in orange.

Procedure

As in Experiments 1 and 2, participants were greeted and given a color vision screening test. A research assistant (RA) logged in and entered a predetermined unique participant ID. The RA directed the participant to read the instructions aloud while the participant completed an example trial.

Unlike the prior experiments, for each trial participants were presented with an exemplar triple that conformed to the rule (see Figure 15). The exemplar triple was unique to this experiment; however, it matched the patterns of the first conforming triple in 14 of the 15 trials (i.e., contained the pattern(s) in the rule and the same distracter pattern(s)). Therefore, having an additional conforming triple provided no essential additional evidence. The program controlled the order the triples were displayed. Specifically, four triples were revealed one at a time in a preset order – depending on the experimental condition – two triples that conformed to the rule and two triples that did not conform to the rule.

The triples were presented to each participant in the same order for all 15 trials consistent with the condition (e.g., in the CNCN condition the participant saw a conforming triple (C), then a nonconforming triple (N), a conforming triple (C), and finally a nonconforming triple (N) for all 15 trials). All conditions presented sufficient

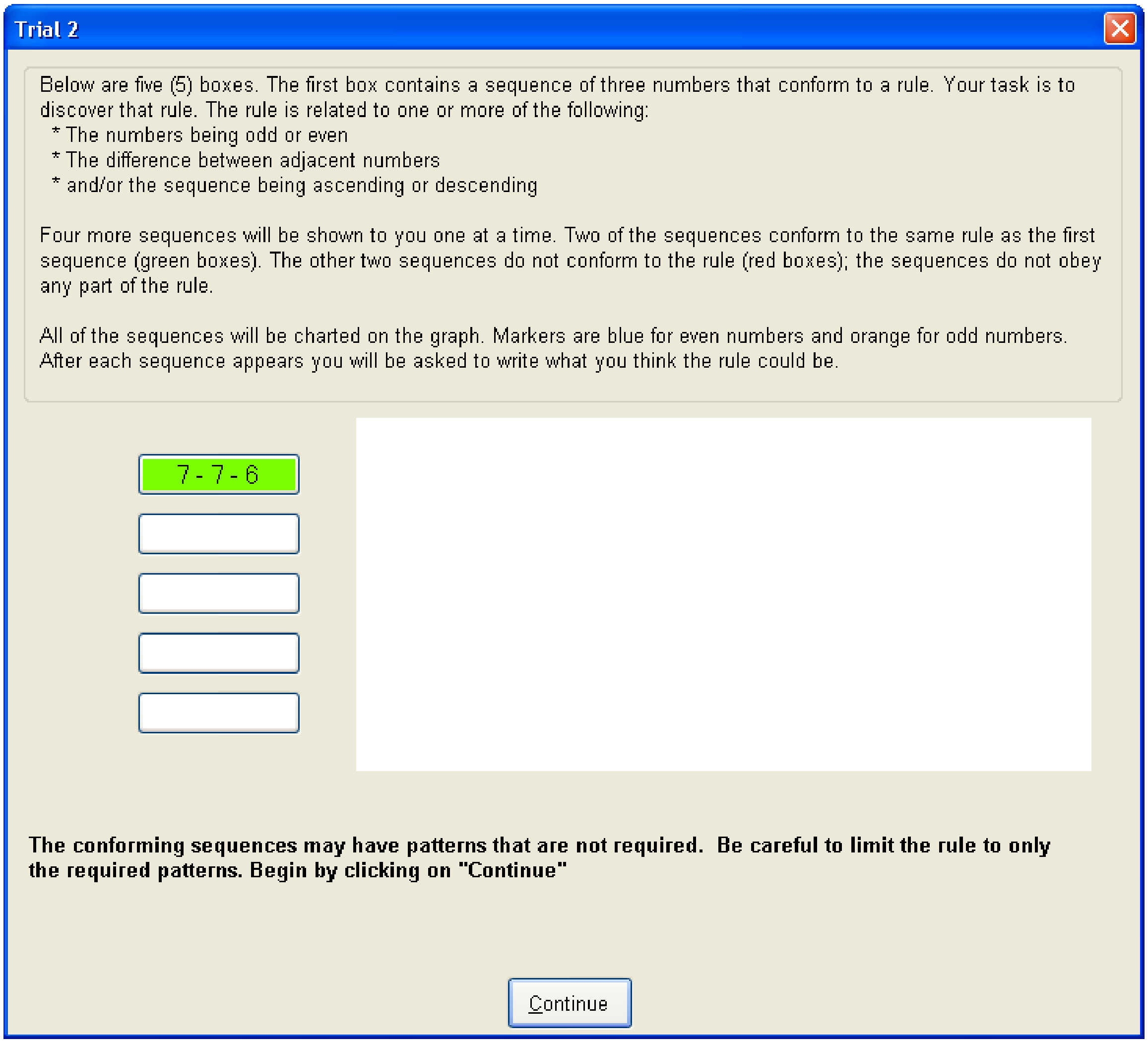

Figure 15. Initial state with exemplar triple

confirming and disconfirming evidence for identifying patterns in the rule and eliminating distracter patterns.

Consistent with Experiment 2, the participants were asked to enter their current hypothesis on what the rule could be after every triple was presented. After five attempts to define the rule, participants were asked to enter the strategy they used to determine the rule (see Figure 16). After entering the strategy, selecting the “Continue” button caused the correct rule to appear. After dismissing the dialog box containing the correct rule, the screen was reset to the next trial.

After 15 trials the participants were presented with a set of questions regarding their strategy (identical to Experiment 2). All participants were also asked “Did the colors and/or graph help? Why or why not?” Participants were presented with the same demographic questions as in Experiments 1 and 2 (see Figure 3).

Measurements

The following measurements were recorded in a database for each trial performed by a participant:

Time. Since the triples were presented in a preset order, the time was recorded for only two participant actions:

- After the rule was entered for each triple
- After the participant had entered the strategy for each trial and selected “Continue”

Rule. The rule measurement was identical to the measurement in Experiment 2, in that participants were instructed to enter what they thought the rule could be after

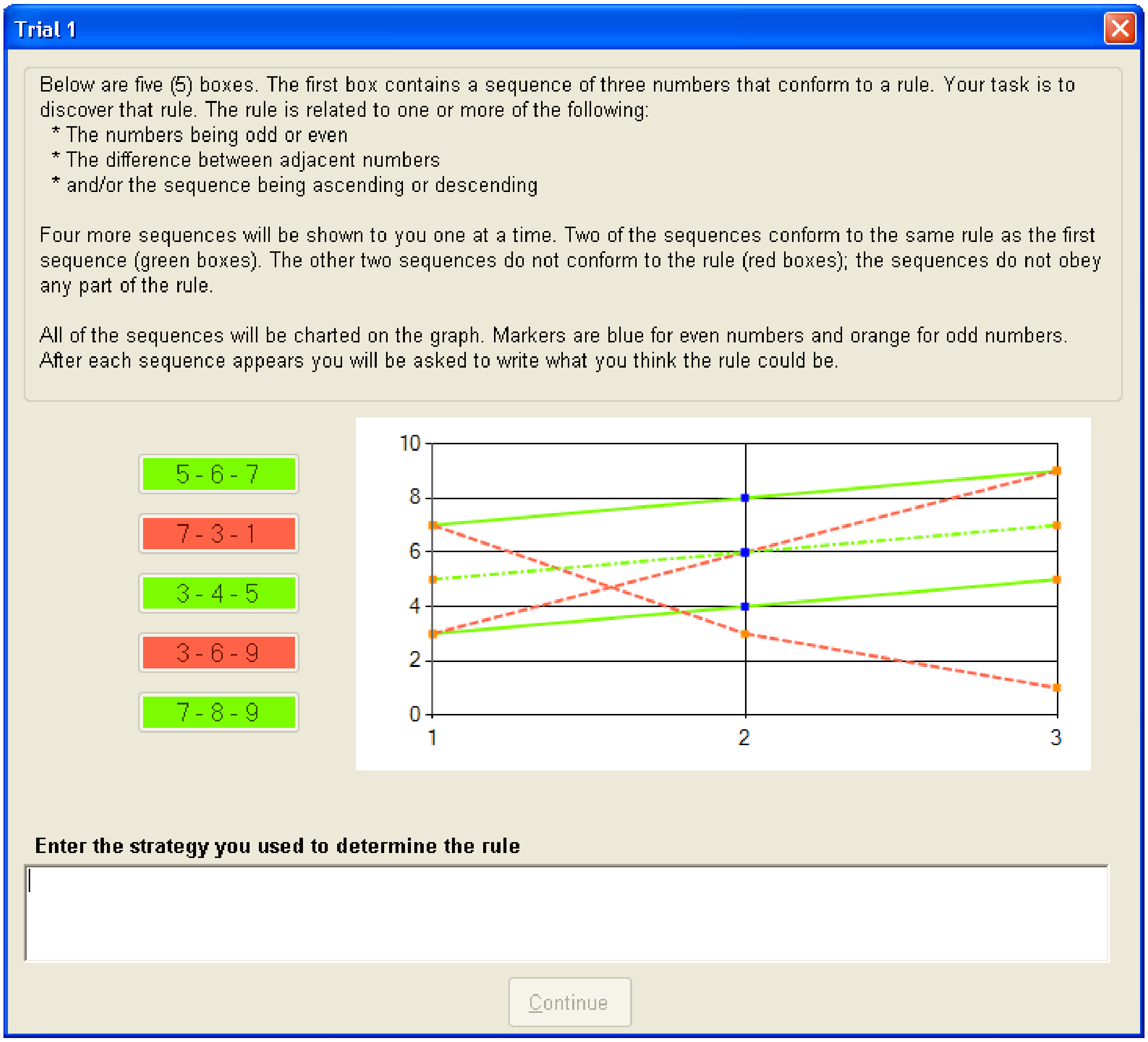

Figure 16. Prompt for strategy for each rule

viewing each triple. The two triples identified as conforming and the two triples identified as nonconforming were identical to the corresponding triples used in Experiments 1 and 2. Therefore, the rules were also identical.

Strategy. Even though the triples were presented in a preset order and no search and selection strategy was required, participants were asked to describe the strategy they used to determine each rule. The example trial displayed the following text for all conditions of the experiments as a description of potential strategies: "There are multiple strategies for determining the rule, one strategy is to consider the given sequence (1-3-5) and think of possible rules for that sequence. You could hypothesize the rule is odd numbers with adjacent numbers differing by 2 and ascending numbers. Looking at other sequences can confirm or disconfirm that hypothesis. Other strategies may be more or less effective; you are encouraged to develop your own strategies."

Consistent with the previous experiments, the purpose of the example text was to describe what was meant by a "strategy" without specifying what the participant's strategy should be. After completing 15 trials the participants were asked if their strategy changed, and if so how did it change.

Data Coding

The results of Experiment 3 were coded and interpreted using methods similar to the previous experiments. Details and exceptions are described below.

Rules

The process for coding the rules was identical to Experiments 1 and 2. All five attempts to define the rule were coded for "hits" and "false positives."

Times

All times were calculated electronically using values captured and stored by the software application. Time to determine the rule was measured from when the participant selected the "Continue" button to start the trial to when the participant completed the last attempt to define the rule.

Data Exclusion

Consistent with Experiment 2, giving participants multiple attempts to define the same rule confused some participants. Thirty-one potential participants failed to understand that there was only one rule per trial. The same percentage (19%) of participants in this experiment misunderstood the task as in Experiment 2. Again, all data from potential participants who demonstrated this misunderstanding were excluded from analyses.

Results

Order

Accuracy

Overall, participants identified an average maximum of 19.74 of the 24 patterns (82.3%) in the rules over the 15 trials, SD = 2.42. Participants identified an average maximum of 15.26 distracter patterns over the 15 trials, SD = 2.67. Participants reduced the distracter patterns from 15.26 to an average of 5.2 by the last attempt.

A one-way ANOVA with the six presentation order conditions as a between subjects factor failed to show a significant difference for identifying the possible 24 patterns in the 15 rules (on the last attempt to define the rule for each trial), F (5, 124) = 1.03, p = .41 (overall M = 18.75 (78.1%), SD = 2.74). There was also no difference in eliminating distracter patterns between presentation order conditions, F (5, 124) = 1.73, p = .13 (overall M = 5.2, SD = 2.98). Based on the last attempt per trial, participants overall accuracy for identifying patterns in the rule while eliminating all the distracter patterns was 8.48 out of 15 trials (56.5%), SD = 2.75.

Time Efficiency

An ANOVA with presentation order as the between subjects factor revealed no difference between presentation orders on total time, F (5, 124) = 1.31, p = .26 (see Table 8). Consistent with Experiment 2, the time to complete the first three trials averaged significantly greater than subsequent trials, F (1, 124) = 22.28, p < .001. There was no interaction of time on the trials and the presentation order, F (5, 124) = .48, p = .79.

Necessary and Sufficient Disconfirming Evidence

All participants were shown necessary and sufficient evidence for determining the rule and eliminating distracter patterns. The attempts to define the rule were used to examine how hypotheses change as more evidence (each triple) was presented. The first exemplar triple conformed to the rule and was sufficient to accurately identify all patterns in the rule. The exemplar triple also contained at least one distracter pattern. Over all 15 trials, on the first attempt participants averaged identifying 16.36 of the 24 (68%) patterns in the rules (minimum = 9; maximum = 24). After viewing all triples the average increased to identifying 18.75 of the 24 (78%) patterns in the rules, (minimum = 12; maximum = 24). A post-hoc analyses found this difference to be significant, t (129) = 9.312, p <.001, (First Attempt, M = 16.36, SD = 3.57; Last Attempt, M = 18.75, SD = 2.76).

Table 8. Mean Total Time for 15 Trials by Order Condition

Total Time

Condition	Mean	Std. Deviation	N
CCNN	1423.80	476.69	21
CNCN	1148.38	355.31	22
CNNC	1183.96	356.74	24
NCCN	1344.09	482.12	20
NCNC	1310.81	390.32	20
NNCC	1285.42	429.73	23
Overall	1278.79	418.88	130

Sufficient versus Insufficient Evidence

Experiments 1 and 2 revealed limitations related to self-directed searches. The effect of sufficient versus insufficient disconfirming evidence was examined by comparing Experiments 2 and 3. Because all of the conforming and nonconforming triples were presented in all trials for Experiment 3, all participants had sufficient evidence to determine the patterns for the rule and to eliminate any distracter patterns. Participants in Experiment 2 selected up to four of the six triples. Only 11% of the searches included selecting all necessary and sufficient evidence – two conforming and two nonconforming triples. None of the participants' searches in Experiment 2 were sufficient for every trial.

Participants in Experiments 3 identified significantly more patterns related to the rule on their last attempt to define the rule $F(1, 170) = 9.02$, $p < .01$, (Exp 2, $M = 17.27$, $SD = 2.90$; Exp 3, $M = 18.75$, $SD = 2.75$). On average, participants in Experiment 3 also identified more distracter patterns before the final attempt, $F(1, 170) = 6.91$, $p < .01$, (Exp 2, $M = 13.95$, $SD = 3.20$; Exp 3, $M = 15.26$, $SD = 2.67$). In addition, participants in Experiment 3 eliminated more distracter patterns by the last attempt, $F(1, 170) = 198.42$, $p < .001$ (Exp 2, $M = 12.52$, $SD = 2.76$; Exp 3, $M = 5.20$, $SD = 2.98$). Experiment 3 also had significantly higher accuracy with trials that eliminated all distracter patterns, $F(1, 170) = 143.02$, $p < .001$ (Exp 2, $M = 3.13$, $SD = 1.60$; Exp 3, $M = 8.48$, $SD = 2.75$).

Having an exemplar triple in Experiment 3 resulted in having three conforming triples and provided participants a potential advantage by giving them one additional attempt to define the rule. A subset of Experiment 3 data was examined to provide a comparison where all participants viewed only two conforming and two nonconforming triples. Specifically, the fourth attempt (of five attempts) to define the rule in Experiment 3 with conditions ending with conforming sequences (i.e., CNNC, NNCC, NCNC) was compared to the last attempt (of four attempts) to define the rule in Experiment 2 trials. This analysis showed participants identified essentially the same number of patterns related to the rule on their fourth attempt to define the rule, $p = .673$, (Exp 2, $M = 17.24$ (71.8%), $SD = 2.90$; Exp 3, $M = 17.55$ (73.1%), $SD = 3.59$). Likewise, participants identified a similar maximum number of distracter patterns before the fourth attempt, $p = .284$, (Exp 2, $M = 13.95$, $SD = 3.20$; Exp 3, $M = 14.56$, $SD = 2.65$). The analysis was biased against the Experiment 3 subset for eliminating distracter patterns since the second conforming pattern was the only way to eliminate one distracter pattern in two of the 15 trials. However, participants in Experiment 3 still eliminated significantly more distracter patterns than participants in Experiment 2 on the fourth attempt, $F(1, 108) = 43.86$, $p < .001$, (Exp 2, $M = 12.52$, $SD = 2.76$; Exp 3, $M = 7.82$, $SD = 4.05$). The Experiment 3 subset also maintained greater accuracy with trials eliminating all distracter patterns, $F(1, 108) = 41.66$, $p < .001$ (Exp 2, $M = 3.13$, $SD = 1.60$; Exp 3, $M = 6.63$, $SD = 3.27$).

Other Factors and Decision Accuracy

A limitation noted in the prior experiments was that the selection of sufficient evidence did not ensure accuracy – given there was no difference in accuracy in Experiment 1 between sufficient and insufficient searches. However, searches in Experiment 1 were overwhelmingly sufficient for determining the rule (80%). Presentation of evidence in Experiment 3 was 100% sufficient for determining the rule. Since overwhelmingly participants in both Experiments 1 and 3 were shown sufficient evidence, comparisons of Experiment 1 and 3 explore factors other than search sufficiency.

One difference between Experiment 1 and 3 was the number of attempts to define the rule. During Experiment 1 the participants selected up to six triples before attempting to define the rule; therefore, participants had only one attempt to define the rule for each trial. During Experiment 3 an exemplar triple and then more four triples were presented in a preset order with an attempt to define the rule after each triple was presented. Participants in Experiment 3 had five attempts to define each rule.

Comparing the attempts to define the rules in Experiment 1 to the first attempt to define the rules in Experiment 3 (both without trial 14) showed no difference in identifying patterns in the rule, $F(1, 208) = .264$, $p = .61$ (Exp 1, $M = 15.67$, $SD = 3.26$; Exp 3, $M = 15.42$, $SD = 3.51$). On the last attempt in Experiment 3 participants eliminated significantly more distracter patterns than participants in Experiment 1, $F(1, 208) = 38.69$, $p < .001$, (Exp 1, $M = 7.0$, $SD = 3.16$; Exp 3, $M = 4.52$, $SD = 3.57$).

Discussion

Some potential limitations of self-directed searches are, the selection of necessary disconfirming evidence cannot be assumed, regardless of the perceived value of disconfirming evidence; the selection of sufficient evidence does not ensure accuracy; however, insufficient selection of disconfirming evidence results in lower accuracy. These limitations raised the question, can presenting sufficient disconfirming evidence reduce confirmation bias and thereby improve performance? A further potential limitation of self-directed searches is that greater selection of partially conforming triples shows ambiguous evidence is considered more valuable than potentially disconfirming evidence. This misperception of value led to the question, can removing irrelevant ambiguous evidence improve performance? Addressing these questions required presenting information, rather than allowing participants' self-directed searches. Presenting information raised concerns of order effects. The consideration of potential order effects are discussed first, then efforts to overcome self-directed search limitations.

Order

Despite predictions of order effects, Experiment 3 did not reveal any differences in accuracy or efficiency based on evidence presentation order. There are multiple possibilities for this lack of effect. For instance, the small number of triples and their short length may not have taxed working memory sufficiently to create differences between conditions; in addition, the triples remained visible after they were presented reducing memory requirements. Another possibility is the task was not complex; participants' familiarity with numbers and patterns may have reduced tendencies to utilize strategies to relieve cognitive strain associated with order (Payne, Bettman, & Johnson, 1993). Task difficulty is not the only factor that contributes to order effects. Simple word sequences can produce order effects. For example, persons described as "intelligent – tall – mean" may be judged more favorably than if described as "mean – tall – intelligent" (see Hogarth & Einhorn, 1992). This word order example illustrates that the response to evidence order may be an affective response. However, for the present study the nondescriptive numbers were unlikely to elicit affective responses, since the numbers did not correspond to trait adjectives for making social judgments of "likableness." Based on these possibilities and perhaps others, the evidence presentation order appears to have had no effect on the hypotheses generated in this experiment, nor on time required to complete trials.

Necessary and Sufficient Disconfirming Evidence

Sufficient versus Insufficient Evidence

One comparison of interest between Experiments 2 and 3 is the effect of evidence sufficiency on accuracy. However, before sufficiency comparisons can be examined, the design similarities and differences should be considered as potential confounding factors: The designs of Experiments 2 and 3 were similar in that both included identifying triples by color. One design difference was that half of the participants in Experiment 2 were presented with a graph, while all of the participants in Experiment 3 were presented with a graph. However, the results of Experiment 2 showed no accuracy differences between the graph and nongraph conditions. Another design difference was participants in Experiment 2 were able to select up to four triples in a self-directed search, versus Experiment 3's preset presentation of the same four triples; however, in both experiments participants attempted to define the rule after each triple. Given the graph conditions' lack of effect on accuracy, comparisons of these two experiments may reflect selection versus presentation of evidence; however, more likely, differences reflect the effects of sufficient versus insufficient disconfirming evidence.

There was a large difference in considering sufficient evidence between the two experiments. Participants' searches in Experiment 2 were overwhelmingly insufficient (89%) to eliminate distracter patterns; furthermore, none of the participant's searches were sufficient to eliminate distracter patterns for every trial. In contrast, all participants in Experiment 3 were presented with sufficient evidence to determine the rule and eliminate distracter patterns for every trial.

Because participants in both Experiment 2 and 3 had sufficient evidence for identifying patterns in the rule, there was no significant difference in identifying those patterns (p = .673). However, even though participants in Experiment 2 and 3 identified a similar maximum number of distracter patterns (p = .284), the lack of sufficient evidence made a significant difference in eliminating distracter patterns from the rule (p <.001). One response to the question, can presenting sufficient disconfirming evidence improve performance, is clearly that having sufficient disconfirming evidence improves performance – as shown by the increase in eliminating distracter patterns from the rule.

Other Factors and Decision Accuracy

Having sufficient evidence is necessary, but having sufficient evidence does not ensure eliminating distracter patterns. Participants in Experiments 1 and 3 were all overwhelmingly shown sufficient evidence to identify patterns in the rule and eliminate distracter patterns. As with the comparison of Experiment 2 to 3, there was no difference between Experiments 1 and 3 in identifying patterns in the rule (p = .61). However, participants in Experiment 3 eliminated significantly more distracter patterns than participants in Experiment 1 after viewing all the evidence (p <.001).

There are multiple potential explanations for the increased accuracy of Experiment 3. One explanation is the difference in design factors between Experiments 1 and 3. First, the wording of the instructions differed slightly; however, the task was identical – to discover the rule related to patterns in the triples. Second, the designs also differed in that half of the participants in Experiment 1 identified triples by color while all participants in Experiment 3 identified triples by color. In addition, none of the participants in Experiment 1 were shown a graph, while all of the participants in Experiment 3 were shown a graph. While these factors had the potential to affect the results of between experiment comparisons, these factors – instructions, color and graph conditions – were unlikely to produce differences since they did not produce any significant differences within Experiments 1 and 2.

Another explanation of the accuracy differences between Experiments 1 and 3 is that presenting sufficient disconfirming evidence rather than self-directed searches possibly improved performance (increased elimination of distracter patterns). At a minimum, comparisons show that presenting sufficient disconfirming evidence is at least as effective as selecting disconfirming evidence. However, simply presenting versus selecting evidence might not be the only factor contributing to differences. To address the question, can removing irrelevant ambiguous evidence improve performance, Experiments 1 and 3 also differed in that Experiment 3 did not present the evidence identified as ambiguous (partially conforming). Another difference between the experiments is that Experiment 3 provided multiple attempts to define the rule. These factors, removal of ambiguous evidence and multiple attempts, are discussed below.

Ambiguous Evidence

Based on the results comparisons of Experiments 1 and 2, one could conclude that the presence of ambiguous evidence is a distraction that reduces the ability to formulate an accurate rule. This provides a potential explanation for the differences between Experiments 1 and 3 for participants eliminating distracter patterns from the rule even when there was sufficient evidence selected or presented in both experiments. A related conclusion is that the presence of ambiguous evidence is especially detrimental since ambiguous evidence is favored over disconfirming evidence thereby significantly reducing the ability to formulate an accurate rule. This overvaluation of ambiguous evidence provides an explanation for the insufficient searches and results of Experiment 2. Selection of higher valued ambiguous evidence reduced selection of disconfirming evidence thereby reducing the elimination of distracter patterns. Thus providing a potential answer of "yes" to the question, can removing irrelevant ambiguous evidence also improve performance?

Multiple Attempts

Another factor emerged when analyzing the result differences between Experiments 1 and 3. Comparisons between experiments and attempts within trials suggest that multiple attempts to define the rule result in better decisions even with no change in sufficient evidence. Specifically, participants in Experiment 1 had only one attempt to define the rule – after selecting up to six triples – while participants in Experiment 3 had five attempts to define the rule for each trial. As previously stated, there was no difference in identifying patterns in the rule on the first attempt between Experiments 1 and 3 (p = .61). This is not too surprising since all evidence for identifying patterns in the rule was available with a single conforming triple. However, identifying patterns in the rule increased significantly from the first attempt to the last in Experiment 3, $t(129) = 9.312$, $p < .001$ (First attempt $M = 16.36$; last attempt $M = 18.75$) – even with no change in sufficient evidence. Multiple attempts to define the rule clearly increased identification of patterns in the rule.

Continued Suboptimal Performance

Removing irrelevant ambiguous evidence and providing multiple attempts to define the rule markedly improved overall performance in Experiment 3. However, not all participants benefited from multiple attempts to define the rule. For 51% of the trials participants identified all patterns in the rule on the first attempt. Leaving 49% of the trials with potential for improvement. However, the overall significant increase between the first and last attempt for identifying patterns in the rule was due to an increase in only 20.2% of the trials. Furthermore, the increase came after offsetting the 4.6% of the trials where participants identified fewer patterns in the rule on the last attempt.

A review of the data for eliminating distracter patterns from the rule also revealed individual differences between subsequent attempts. During Experiment 3, after viewing the first triple, subsequent triples provided disconfirming evidence for eliminating distracter patterns. At least one distracter pattern was identified on the first attempt in 72% of the trials. In 43.4% of the trials participants did not eliminate any distracter patterns between the first and last attempts to define the rule. For a nearly equal amount (43.9%) of trials participants eliminated one distracter pattern, and for 7.7% of trials participants eliminated two distracter patterns between the first and last attempt. In the remaining 5.1% of trials participants added at least one distracter pattern between the first and the last attempts to identify the rule. In summary, despite providing disconfirming evidence, for 49% of trials no distracter patterns were eliminated between the first and last attempts to define the rule. Why did multiple attempts help some participants and not others? What other factors should be considered in efforts to improve decision support? Consideration of participants' strategies offers some insight to these questions.

Strategies

At the end of the 15 trials participants were asked if their strategy changed and why. If their strategy did change they were asked to explain how it changed. The majority of participants (91, 70%) reported a change in strategy; 32 participants (24.6%) reported no change in strategy. Seven participants (5.4%) responded with "not sure." Some participant's explanation of their strategies were nonspecific responses such as, "It changed based on what went wrong in previous tasks." A characteristic of the associative system is that the person "is conscious only of the result of the computation, not the process" (Sloman, 1996). An associative process correlates to some participants' description of their strategy exemplified by, "Most patterns I could just look at and see the pattern or what the numbers were doing."

Elimination of distracter patterns requires deliberation, analysis and verification. Some participants articulated a more insightful awareness of a rule-based strategy, typified by, "I started including every possible part of the rule on the first green, and then eliminating as the other sequences were shown" and "All I did was come up with all possible rules with the first green box and used the other four boxes to eliminate any rule that did not apply." This Find All Then Eliminate (FATE) strategy was explicitly expressed by 15 participants (11.5%). Participants who expressed their strategies as a rule-based FATE strategy developed more accurate rules than other participants, as shown by a MANOVA analysis considering the patterns identified on the first attempt and the distracter patterns remaining on the last attempt, $F\,(2, 127) = 3.46$, $p = .03$. The FATE strategy resulted in eliminating more distracter patterns from the rule by the last attempt, $F\,(1, 128) = 5.99$, $p = .02$ (FATE, M=3.67; All others, $M = 5.43$). There was no difference in identifying the 24 patterns in the 15 rules on the first attempt, F (1, 128) = .03, p = .86 (FATE, M=16.2 (67.5%); All others, M = 16.4 (68.3%)).

These results raised the question that if participants identify more patterns (both patterns in the rule and distracter patterns) on the first attempt are they more likely to identify patterns in the rule and eliminate more distracter patterns after all the evidence has been presented? One hypothesis is that identifying all possible patterns in the first triple frees up cognitive resources and supports a switch to rule-based processing that uses the rest of the triples to eliminate distracter patterns. A correlation analysis supports this hypothesis in that identifying patterns in the rule correlates positively with identifying distracter patterns on the first attempt, $\tau_b = .45$, $p < .001$. Further, that more distracter patterns are eliminated if found early is shown by the correlation between the number of distracter patterns found on the first attempt and the increase in accuracy with all distracter patterns eliminated by the last attempt τ_b (130) = .15, p = .02.

Summary

The results of Experiment 3, in conjunction with comparisons to Experiments 1 and 2, suggest affirmative answers: Presenting sufficient disconfirming evidence does have the potential to reduce confirmation bias and improve performance; in addition, removing (or not presenting) ambiguous evidence may also improve decision making performance.

These answers do not provide a complete solution for supporting overcoming confirmation bias. Furthermore, the hypothesis that identifying all possible patterns in the first triple frees up cognitive resources and supports a switch to rule-based processing provides only a potential explanation. The question remains of how to promote a strategy like the FATE strategy to facilitate the switch from associative to rule-based processing. Providing or requiring multiple attempts may support a switch to more rule-based processing, but, in this research did not ensure the switch (since providing for multiple attempts did not ensure greater accuracy for all participants). Perhaps rule-based processing requires more effort than some participants were willing to expend or participants lacked the understanding or ability to employ an effective strategy.

GENERAL DISCUSSION

This research explored data presentation and human cognition with the objective of improving electronic decision support systems. Electronic decision support remains a concern because in spite of 60 years of effort, electronic computing has failed to reliably replace human cognition in complex domains. Some factors in this failure are that suboptimal properties of the data and complexities of the domain often require human interpretation and intervention. Human interpretation relies on experience, values, intuition, insight and learning, which can lead to shortcuts or heuristics. Heuristics in the correct context can be economical and effective in solving many problems. However, cognitive biases are failed heuristics that can lead to errors. Biases all share the elements of structuring incorrect or inappropriate models and hypotheses and/or insufficient consideration of the data. Most if not all biases can be linked to confirmation bias – which is manifested by searching for only confirming data.

The present research explored de-biasing techniques to shift cognitive processing from an automatic associative mode to a more deliberate, conscious rule-based mode. The general question was when and how to facilitate the shift in the context of decision support. The experiments are summarized in the next section, followed by sections on implications for decision support systems, limitations of the study and future directions.

Summary of Experiments and Findings

Experiment 1

All three experiments were adaptations of the Wason 2-4-6 task. The first experimental design addressed two questions: Can increased salience of data characteristics reduce confirmation bias and thereby improve performance; and, can instructions to structure appropriate hypotheses and consider data sufficiently, reduce confirmation bias and thereby improve performance?

Experiment 1 results were inconclusive on the question of using increased salience of data characteristics to reduce confirmation bias. With an unrestricted search participants typically selected all available evidence, and neither increased salience of data characteristics (using color versus text) nor instructions (to consider disconfirming evidence versus generating multiple hypotheses) affected their search sufficiency, efficiency or decision accuracy. Moreover, instructions to consider disconfirming evidence did not increase the perceived value of disconfirming evidence.

Experiment 2

Experiment 2's design limited the ability to select data, to manipulate scarcity, and facilitated judging participants' perceived value of the data. The experiment also examined the effect of increased salience of data patterns by comparing charting data on a graph versus no graph. Charting the data on a graph did not increase selection of disconfirming data, nor did it change decision accuracy. The only measured effects of the graph were an increase in the amount of time participants required to complete the study and an increased failure to eliminate ascending or descending distracter patterns from the rule.

The results of restricting the search to manipulate scarcity does provide insight into what evidence was considered most valuable. Not surprisingly, conforming data were considered the most valuable (all conforming data were selected in 98% of trials). Interestingly, ambiguous evidence was considered more valuable than potentially disconfirming evidence – all ambiguous evidence was selected nearly twice as often as disconfirming data. Even when disconfirming evidence was perceived as valuable, participants still selected more ambiguous evidence than disconfirming evidence.

The results of Experiments 1 and 2 led to four conclusions on potential limitations of self-directed searches: 1) The selection of necessary disconfirming evidence cannot be assumed in self-directed searches, regardless of the perceived value of disconfirming evidence. 2) The selection of sufficient evidence does not ensure accuracy regardless of the perceived value; however, 3) insufficient disconfirming evidence does result in lower accuracy. The corollary is the presence of disconfirming evidence does result in increased accuracy. 4) Selecting ambiguous evidence may be more compelling than selecting disconfirming evidence. The selection of ambiguous evidence over truly disconfirming evidence can be detrimental. The preference for ambiguous evidence over disconfirming evidence is shown regardless of the stated value of disconfirming evidence, perhaps because ambiguous evidence is perceived as disconfirming.

Experiment 3

In response to the limitations of self-directed searches, two questions were considered in Experiment 3: 1) Can presenting sufficient disconfirming evidence reduce confirmation bias and thereby improve performance; and 2) can removing irrelevant ambiguous evidence also improve performance? Presenting evidence rather than allowing self-directed searches raised concerns and suggested hypotheses for order effects.

Experiment 3 failed to produce the hypothesized order effects. However, the results of Experiment 3 support the conclusion that having sufficient disconfirming evidence reduces confirmation bias – resulting in improved performance. Further, that presenting sufficient disconfirming evidence is at least as effective as selecting disconfirming evidence.

In response to the second question, can removing irrelevant ambiguous evidence improve performance, the increased accuracy in Experiment 3 compared to Experiment 1 may have resulted from removing ambiguous evidence or from requiring multiple attempts to define the rule or a combination of both factors.

Regardless of the condition, accuracy increased for some participants but not others. One explanation may be the strategy employed. A find all possible rules then eliminate distracter patterns appeared to be the most effective strategy. This suggests that identifying all possible patterns or diagnoses frees up cognitive resources and supports a switch to rule-based processing that uses subsequent evidence to eliminate distracters and refine the conclusions. The FATE strategy may explain differences between performances in terms of associative and rule-based processing, however, the question remains of how to promote the switch from associative to rule-based processing.

Implications for Decision Support Systems

Reliance on Increased Salience

Increasing the salience of data patterns does not ensure increased search efficiency, sufficiency or decision accuracy. The design of Decision Support Systems

(DSS) should not rely exclusively on increased salience to improve search performance. Prior research (Kramer, 2010) showed that manipulating salience of evidence – by size and/or position – may increase search efficiency but generally fails to increase search sufficiency or decision accuracy. This study is consistent in that increased salience by identifying patterns related to the task does not ensure increased search sufficiency or decision accuracy. In fact, increasing salience of evidence through graphs can decrease efficiency and reduce accuracy. The salience of the ascending and descending patterns not only failed to result in greater efficiency, but the potential support of associative processing possibly hindered a switch to rule-based processing.

Instructions as a Method to Debias Search

The results of Experiment 1 suggest that instructions to consider disconfirming evidence were ineffective and therefore did not achieve the desired de-biasing effect. There are likely multiple reasons for the instruction's lack of effectiveness. One reason may be that implementing the instructions to consider multiple hypotheses and/or disconfirming data requires domain knowledge (see Willingham, 2008). Therefore, in addition to instructing the user to consider multiple perspectives, designers of a DSS should carefully consider whether the user has the prior domain knowledge. For example, it is logical that a physician or any other domain expert must have knowledge of alternative hypotheses, expected measurements and observations before performing an adequate differential diagnosis and eliminate diagnoses that do not conform to the evidence. That designers of a DSS should carefully consider whether the user has the prior domain knowledge seems like an obvious conclusion; nevertheless, an inappropriate assumption of domain knowledge may provide partial explanation for the failure of some decision support systems.

The findings of the present research reflect the mixed efficacy findings of other research exploring instructions and training to de-bias decision making (see Lilienfeld et al., 2009). Facilitating the switch to rule-based processing may require directed and substantial effort since associative cognition is typified by the person being "conscious only of the result of the computation, not the process" (Sloman, 1996).

Perceived Value and Search Sufficiency

Designers of DSS should ensure that disconfirming evidence not only be available, but presented to the user. Experiment 2 results reflect that when resources are limited, searches are rarely sufficient even when disconfirming evidence has perceived value. However, regardless of the perceived value there is a significant benefit to having sufficient evidence available.

The concept of presenting disconfirming evidence along with any confirming search results is supported by Kayhan's study (2013) where participants received recommendations to view disconfirming evidence ("recommendation" condition) or the search results incorporated disconfirming evidence even if only confirming evidence was sought ("incorporation technique" condition). In the "recommendation" condition no participants viewed the disconfirming evidence and 75% indicated a valid relationship with the hypothesis. In the "incorporation technique" 75% of the participants viewed the disconfirming evidence and 75% disagreed with the hypothesis. Confirming and disconfirming evidence are necessary to form a hypothesis and data consideration sufficient to eliminate errors. An important conclusion is that even if not requested, presenting disconfirming evidence can affect the decision.

In addition to presenting disconfirming evidence, removal of ambiguous evidence that does not offer value may be beneficial, especially since the ambiguous evidence may be perceived as more valuable than disconfirming evidence and therefore be given greater and unwarranted consideration.

Multiple Attempts and Rule-Based Processing

Given sufficient domain knowledge, developing hypotheses may become an associative task; however, eliminating distracter models often requires a switch to more deliberate rule-based processing. Therefore, a DSS might increase accuracy by supporting generation of hypotheses and promoting a switch to rule-based processing through multiple attempts to reach conclusions. This correlates to recommendations by Spengler et al. (1995) for clinical psychologists to slow down decision making, emphasizing the importance of making tentative judgments which are “repeatedly subjected to rejection.” The challenge for a DSS would be to address the situations where users have reached closure and would likely be frustrated with a system that appears to arbitrarily slow them down.

Limitations

Decision making is a complex cognitive activity and the study of decision making encompasses many domains with varied human factors and data sets. Even with the limited focus of the present study there are known and likely unknown confounding and missing factors. A few of the obvious limitations of this research are the lack of consideration of individual differences, the design of the task, and naiveté of the participants. These limitations are expounded below.

Individual differences were not measured in the present work. Multiple individual differences potentially affected the identification of patterns and elimination of distracter patterns. Individual differences based on factors such as, working memory capability, personality, values, motivation, knowledge, skill and experience, all potentially influence cognition and decision making (see Daneman & Carpenter, 1980; Kanai & Rees, 2011; Klein, Phillips, Rall, Peluso, & Hoffman, 2007; Levin, Huneke, & Jasper, 2000; Venkatesh & Morris, 2000). Also, the need for cognitive closure is a trait that varies between individuals and situations, and is marked by a quick final decision with low tolerance for uncertainty and ambiguity (see Choi, Koo, Choi, & Auh, 2008). In addition, it may be that some individuals have more difficulty overcoming the associative recognition of patterns (Sloman, 1996) and are therefore less successful in eliminating patterns.

There may be factors in the study itself that limit the interpretation and generalizability of the results. The abstract study design may not generalize to real-life decision making contexts. Specifically, the simplistic rules in this study may not correlate to the complexity of environments with conflicting, ambiguous and excessive data. However, given the overall less than optimal performance in this study, it may be assumed that increased complexity is not necessary to reveal errors arising from confirmation bias.

Another limitation of the present study is that participants were undergraduate students with no expected expertise or training in the task. Even with the seemingly simple task in this study, it is possible participants lacked the knowledge to recognize and eliminate distracter patterns. Domain experts may react differently than novices to increased salience, instructions, and/or familiar charted data.

Future Directions

Future work should consider applying the findings of this study to the design, implementation and validation of decision support systems. Studies should be done with users who have domain knowledge. In addition, given the importance of disconfirming evidence, more exploration is required for when and how to present unsought disconfirming evidence. Moreover, given the potential for distraction and overrated value, more attention should be paid to the effects of ambiguous evidence. Further consideration should also be given to the trade-offs between efficiency through associative processing and when and how to facilitate a switch to rule-based processing. Specifically, when the circumstances warrant, explore facilitating generation of multiple hypotheses which are subsequently subjected to repeated possibilities for rejection.

REFERENCES

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review, 31*(2-3), 79-85. doi: http://dx.doi.org/10.1016/j.dr.2011.07.002

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and the order of information. *Medical Decision Making, 18*(4), 412-417.

Booth, A., Carroll, C., Ilott, I., Low, L. L., & Cooper, K. (2013). Desperately seeking dissonance: Identifying the disconfirming case in qualitative evidence synthesis. *Qualitative Health Research, 23*(1), 126-141. doi: 10.1177/1049732312466295

Chen, W.-T., & Styler., W. (2013). *Anafora: A web-based general purpose annotation tool.* Paper presented at the NAACL HLT 2013, Atlanta, GA.

Cherubini, P., Castelvecchio, E., & Cherubini, A. M. (2005). Generation of hypotheses in Wason's 2-4-6 task: An information theory approach. *Quarterly Journal of Experimental Psychology: Section A: Human Experimental Psychology, 58*(2), 309-332.

Choi, J. A., Koo, M., Choi, I., & Auh, S. (2008). Need for cognitive closure and information search strategy. *Psychology & Marketing, 25*(11), 1027-1042.

Croskerry, P. (2002). Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine, 9*(11), 1184-1204. doi: 10.1197/aemj.9.11.1184

Cunnington, J. P., Turnbull, J. M., Regehr, G., Marriott, M., & Norman, G. R. (1997). The effect of presentation order in clinical decision making. *Academic Medicine, 72*(10), S40-42.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi: http://dx.doi.org/10.1016/S0022-5371(80)90312-6

Dick, R. S., Steen, E. B., & Detmer, D. E. (Eds.). (1997). *The computer-based patient record: An essential technology for health care* (Revised ed.). Washington, D.C.: The National Academies Press.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York City: Teachers College, Columbia University

Englund, M. P., & Hellström, Å. (2012). If you have a choice, you have trouble: Stimulus valence modulates presentation-order effects in preference judgment. *Journal of Behavioral Decision Making, 25*(1), 82-94. doi: 10.1002/bdm.714

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459. doi: http://dx.doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2-3), 86-102. doi: http://dx.doi.org/10.1016/j.dr.2011.07.007

Few, S. (2006). *Information dashboard design: The effective visual communication of data* (First ed.). Sebastopol, CA: O'Reilly Media Inc.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Penguin.

Gorman, M. E., & Gorman, M. E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36A*(4), 629-648.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology, 40*(3), 374-383.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*(6), 1069-1086. doi: 10.1037/0022-3514.69.6.1069

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*(1), 1-55. doi: http://dx.doi.org/10.1016/0010-0285(92)90002-J

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundation of cognitive systems engineering*. Boca Raton, FL: CRC Press.

Jarvenpaa, S. L., Dickson, G. W., & DeSanctis, G. (1985). Methodological issues in experimental IS research: Experiences and recommendations. *MIS Quarterly, 2*(9), 141-156.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Macmillan.

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience, 12*(4), 231-242. doi: 10.1038/nrn3000

Kayhan, V. O. (2013). Seeking health information on the web: Positive hypothesis testing. *International Journal of Medical Informatics, 82*(4), 268-275.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211-228.

Klein, G., Phillips, J. K., Rall, E. L., Peluso, D. A., & Hoffman, R. R. (2007). A data-frame theory of sensemaking. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp. 113-155). Mahwah, NJ Lawrence Erlbaum Associates Publishers.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*(3), 185-225.

Kramer, H. S. (2010). *The cost of saliency: How salience interacts with relevance to support or impede information search* (Master of Science), University of Utah, Salt Lake City, UT.

Lange, N. D., Thomas, R. P., & Davelaar, E. J. (2012). Temporal dynamics of hypothesis generation: The influences of data serial order, data consistency, and elicitation timing. *Frontiers in Psychology, 3*.

Lapinski, A.-L. S. (2009). A strategy for uncertainty visualization design (Vol. DRDC-ATLANTIC-TM-2009-151). Dartmouth (Canada): Defence Research and Development Atlantic.

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion/exclusion effects. *Organizational Behavior and Human Decision Processes, 82*(2), 171-193.

Liddy, E. D. (2001). Natural language processing *Encyclopedia of Library and Information Science* (2nd ed.). New York, NY: Marcel Decker, Inc.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4*, 390-398. doi: 10.1111/j.1745-6924.2009.01144.x

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231-1243. doi: 10.1037/0022-3514.47.6.1231

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science, 32*(3), 139-160.

Mumma, G. H., & Wilson, S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of Clinical Psychology, 51*(6), 841-853.

Nickerson, R. A. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *The Visual Computer, 13*(8), 370-390. doi: 10.1007/s003710050111

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. New York, NY US: Oxford University Press.

Puvathingal, B. J., & Hantula, D. A. (2012). Revisiting the psychology of intelligence analysis: From rational actors to adaptive thinkers. *American Psychologist, 67*(3), 199-210.

Rebitschek, F. G., Scholz, A., Bocklisch, F., Krems, J. F., & Jahn, G. (2012). Order effects in diagnostic reasoning with four candidate hypotheses. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 905-911). Austin, TX: Cognitive Science Society.

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association, 8*(6).

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138.

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research, 6*(1), 1-10.

Sittig, D. F., Wright, A., Simonaitis, L., Carpenter, J. D., Allen, G. O., Doebbeling, B. N., . . . Middleton, B. (2010). The state of the art in clinical knowledge management: An inventory of tools and techniques. *International Journal of Medical Informatics 79*(1), 44-57. doi: 10.1016/j.ijmedinf.2009.09.003

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

South, B. R., Shen, S., Leng, J., Forbush, T. B., DuVall, S. L., & Chapman, W. W. (2012). *A prototype tool set to support machine-assisted annotation.* Paper presented at the 2012 Workshop on Biomedical Natural Language Processing, Montreal, Canada.

Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice and research *The Counseling Psychologist 23*. doi: 10.1177/0011000095233009

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2-3), 103-118. doi: http://dx.doi.org/10.1016/j.dr.2011.07.003

Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Brain Research, 1287*(0), 118-125. doi: http://dx.doi.org/10.1016/j.brainres.2009.06.080

Tufte, E. R. (2000). *Visual explanations* (4th ed.). Cheshire, CT: Graphics Press.

Turban, E., & Aronson, J. E. (2000). *Decision support systems and intelligent systems*. Upper Saddle River, NJ: Prentice Hall PTR

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Vallée-Tourangeau, F., & Payton, T. (2008). Graphical representation fosters discovery in the 2-4-6 task. *Quarterly Journal of Experimental Psychology, 61*(4), 625-640.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115-139.

Wainer, H. (2005). *Graphic discovery: A trout in the milk and other visual adventures* Princeton, NJ: Princeton University Press.

Walker, L., Thibaut, J., & Andreoli, V. (1972). Order of presentation at trial. *The Yale Law Journal, 82*(2), 216-226. doi: 10.2307/795112

Wang, H., Zhang, J., & Johnson, T. R. (2000). Human belief revision and the order effect *Twenty-second Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*(3), 129 - 140.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology, 61*(4).

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review, 109*(4), 21-32.

Wright, A., Sittig, D. F., Ash, J. S., Feblowitz, J., Meltzer, S., McMullen, C., . . . Middleton, B. (2011). Development and evaluation of a comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed electronic health record systems. *Journal of the American Medical Informatics Association*(18), 232-242. doi: 10.1136/amiajnl-2011-000113

Zuk, T., & Carpendale, S. (2006). *Theoretical analysis of uncertainty visualizations.* Paper presented at the SPIE 6060, Visualization and Data Analysis doi:10.1117/12.643631.

Zuk, T., & Carpendale, S. (2007). *Visualization of uncertainty and reasoning*. Paper presented at the 7th International Symposium on Smart Graphics, Kyoto, Japan.