TRAINING AWAY ANCHORING IN A WEIGHTED CENTROID JUDGMENT

TASK

by

SHAW LIONEL KETELS

B. A., University of Colorado 2000

M. A., University of Colorado, 2009

A dissertation submitted to the Faculty of the Graduate School of the University of Colorado in partial fulfillment of the requirement for the degree of Doctor of Philosophy Department of Psychology and Neuroscience 2013 This dissertation entitled: Training Away Anchoring in a Weighted Centroid Judgment Task written by Shaw Lionel Ketels has been approved for the Department of Psychology and Neuroscience

	Alice F. Healy	
Signature		
	Matt Jones	
Signature		
	Michael Mozer	
Signature		_
	Lyle E. Bourne, Jr.	
Signature		
	Tamara Sumner	
Signature		

Date May 29, 2013

The final copy of this thesis has been examined by the signatories, and we Find that both the content and the form meet acceptable presentation standards Of scholarly work in the above mentioned discipline.

IRB protocol # <u>11-0546</u>

Ketels, Shaw Lionel (Joint Ph.D., Cognitive Psychology & Cognitive Science) Training Away Anchoring in a Weighted Centroid Judgment Task Dissertation directed by College Professor of Distinction Alice F. Healy

Initial impressions are lasting, and thus initial misunderstandings can hinder subsequent performance in many domains. In previous work we described evidence of the anchoring bias in a centroid judgment task involving sequentially arriving targets, varying in spatial location. In decisions based on sequentially arriving pieces of information, the anchoring bias has been suggested to lead to order effects, or a greater influence at items at a particular serial position on decisions. This dissertation describes five experiments using this centroid judgment task. The first experiment examines the effect of complexity on primacy and recency in this simple decision. The second, third, and fourth experiments explore declarative and nondeclarative approaches to debiasing these anchoring effects, without success. The fifth experiment explores the effect of articulatory suppression on bias in the centroid judgment. Results of the final experiment suggested that anchoring might involve effortful, explicit processing to a greater extent than is often suggested, because articulatory suppression eliminated any demonstrated anchoring bias in the decision.

CONTENTS

CHAPTER			
I.	INTRODUCTION1		
	Dual-process Framings of Cognition4		
	Anchoring, Primacy, & Recency9		
	Debiasing12		
II.	OVERVIEW OF EXPERIMENTS 15		
III.	EXPERIMENT 1		
IV.	EXPERIMENT 2		
V.	EXPERIMENT 3		
VI.	EXPERIMENT 4 50		
VII.	EXPERIMENT 5		
VIII.	CONDITIONS WITHOUT DEBIASING		
IX.	GENERAL DISCUSSION		
	The Effect of Practice		
	Theoretical Implications70		
	Conclusion73		
BIBLIOGRAPHY75			
APPENDIX			
A.	PRELIMINARY EXPERIMENTS		
B.	CONTRAST OF PRELIMINARY AND PRESENT EXPERIMENTS		

TABLES

Table

1.	<i>Comparison of the conditions of the five experiments</i>	
11		

FIGURES

1.	The feedback screen	. 15
2.	Calculation of the optimal deployment location	. 20
3.	Experiment 1 - Deployment accuracy	. 31
4.	Experiment 1 - Deployment bias	. 32
5.	Experiment 1 – Recall accuracy	. 34
6.	Experiment 2 - Deployment accuracy	. 38
7.	Experiment 2 - Deployment bias	. 40
8.	Experiment 2 – Recall accuracy	.41
9.	Experiment 3 - Deployment accuracy	. 45
10.	Experiment 3 - Deployment bias	. 47
11.	Experiment 3 – Recall accuracy	. 48
12.	Experiment 4 - Deployment accuracy	. 52
13.	Experiment 4 - Deployment bias	.54
14.	Experiment 4 – Recall accuracy	. 56
15.	Experiment 5 - Deployment accuracy	. 59
16.	Experiment 5 - Deployment bias	.60
17.	No debiasing conditions – Decision accuracy	.61
18.	Preliminary Experiment 1 – End analysis	.86
19.	Preliminary Experiment 1 – Deployment accuracy	. 88
20.	Preliminary Experiment 1 – Deployment accuracy, free vs. serial	. 89
21.	Preliminary Experiment 1 – End bias, free vs. serial	.90

22.	Preliminary Experiment 1 – End analysis, week 1 vs. week 2
23.	Experiment 2 – Deployment accuracy92
24.	Experiment 2 – End Analysis

vii

Chapter I Introduction

Each such impression, be it of a theoretical issue, another person, or a social organization, grows and changes over the course of time. At any point in time, therefore, the current impression looks both forward and back (p.144). –Norman Anderson, 1981

There are a number of real-world situations in which information is processed serially, spaced over varying timescales. In certain examples, such as medical and military situations, departures from optimality can result in injury and/or death. Simple decisions based on sequentially arriving information have thus been investigated extensively, across multiple lines of literature (Anderson, 1981; Cook & Smallman, 2008; Hogarth & Einhorn, 1992). Investigators have described a number of item characteristics that influence these decisions. In the majority of cases in various literatures, earlier presented items influence subsequent decisions to a greater extent than do items presented later in lists, a primacy or anchoring effect. However, in other cases the last items presented also exert a large influence, a recency effect. The anchoring bias, arising from capacity limited cognitive constructs such as attention and working memory, has been blamed for these effects – processing limitations lead to a reliance on incomplete information across contexts (e.g., Tversky & Kahneman, 1974; Hastie & Park, 1986; Hogarth & Einhorn, 1992).

Although anchoring is a ubiquitous finding, it has been an elusive target for explanation. Investigators usually use some conceptualization involving two general cognitive systems as a framework within which to discuss these effects. This dual conceptualization should be relatively unsurprising for two reasons. First, a general distinction between two forms of learning from, and acting on, the environment has been philosophically, experimentally, and colloquially ubiquitous for thousands of years (e.g., Baars, 1988; Jaynes, 1976; Neely, 1977; Polanyi, 1958; Schacter, 1992; Sloman, 1996). Specific instantiations vary widely, and so these characterizations have been referred to generally as "dual-process" theories of cognition (e.g., Evans, 2008; Frankish & Evans, 2009). Second, Tversky and Kahnemann's original discussion of anchoring and other heuristics employed a dual-process framing, contrasting "intuitive" and "rational" judgments (Tversky & Kahnemann, 1974). Their usage created a salient anchor with a large influence on all subsequent investigators interested in anchoring.

Anchoring and serial position effects in memory have both been popular empirical targets since their introduction, but despite numerous commonalities, the two literatures have remained largely independent of one another. "Serial position effects" refer to better recall for items that occur at certain serial positions in sequences of presented information - those serial positions nearer the beginning and ends of lists are remembered better than those in the middle. Elaborated versions of Tversky and Kahnemann's (1974) original account continue to be sufficient explanations for most investigators interested in anchoring (e.g., Hogarth & Einhorn, 1992; LeBoeuf & Shafir, 2009). Tversky and Kahnemann described an "anchoring-andadjustment" process, whereby initially presented information serves as an "anchor," with each new piece of information that must be considered in making the decision leading to an "adjustment" (1974). However, people are conservative in these adjustments, and so the final impression or decision does not deviate far from the anchor. The generalizability of the anchoring-and-adjustment account has been questioned, largely based on the assertion that an anchoring-and-adjustment mechanism can only account for anchoring that results from anchors whose values are extreme with respect to the total relevant range of decision values (e.g., Mussweiler & Strack, 1999). Popular alternative (or supplemental) accounts of anchoring involve activation - whichever information is most active influences decisions to the greatest

2

extent (e.g., Chapman & Johnson, 1999; Mussweiler & Strack, 1999). Examples include explanations of anchoring based on either numeric or semantic priming.

"Debiasing" refers to influencing subjects to cease exhibiting decisional biases, and thus instead execute decisions more rationally. The severe real-world consequences of biases in certain decisions make debiasing an important topic in the human factors literature (e.g., Arnott, 2006; Cook & Smallman, 2008), and it has also remained a popular topic in the judgment and decision making literature, where the anchoring heuristic was originally described (e.g., Chapman & Johnson, 1999). Debiasing can be cast as training people to exert explicit, intentional, or type 2 (t2) control over implicit, automatic, or type 1 (t1) processes (e.g. Wilson, Houston, Etling, & Brekke, 1996). T1 processes are difficult to modify, because they are "process-opaque", or because we lack conscious access to them. Therefore although there are some examples of success, most debiasing efforts have been met with failure (Cook & Smallman, 2009; Furnham & Boo, 2011; LeBoeuf & Shafir, 2009).

We have examined anchoring in a paradigm in which we ask subjects to make a centroid judgment based on seven sequentially presented spatial locations (Ketels, Healy, Wickens, Buck-Gengler, & Bourne, 2011). "Centroid judgment" refers to choosing a location that is central to a set of other locations, or choosing a point that minimizes the distance to each of a set of other points. In the first few iterations of the paradigm we found a consistent primacy effect, but no recency effect, in the decision. That is, subjects chose centroids that were closer to spatial positions of the first two items in each sequence than the items at any other serial position, with chosen centroids similarly distant from the other five stimulus locations. As described in Appendix A, we also found that instructional debiasing had an effect in this paradigm – within the context of our cover story, we instructed subjects to try to choose centroid locations that were

3

closer to stimuli appearing later in each sequence. Although these debiasing instructions did not alter the inordinate influence of the first item presented in each sequence on decisions, these instructions led to an increased influence of the last few items on centroid judgments.

Paradigms in the anchoring literature usually rely on semantic information, thus the first experiment reported here examined whether subjects making simple perceptual decisions such as centroid judgments exhibit the same biases as subjects making other kinds of judgments where anchoring consistently obtains. Thus Experiment 1 was intended to verify the existence of the expected pattern of bias across three levels of stimulus complexity in this paradigm. Experiments 2-5 examine strategies intended to attenuate the inordinate influence of the initially presented item on centroid judgments, within the context of the highest stimulus complexity from Experiment 1. We used the highest level of stimulus complexity here to make contact with as many extant debiasing paradigms as possible, because the basic version of the task here is much less complex than other empirical contexts within which these questions have been examined.

Dual-process framings of cognition

Since the original study by Tversky and Kahnemann (1974), the application of heuristics has been described as "automatic" or "intuitive", and thus difficult to modify. However, the opposite argument is also often implied - the difficulty in modifying the application of heuristics is offered as evidence of the automatic and/or intuitive nature of the biases that result from their application. However, in the case of anchoring and adjustment, adjustment processes have been at times described as "explicit," "effortful," and/or "controlled" processes (e.g., Epley & Gilovich, 2005; Furnham & Boo, 2011). Nonetheless, the anchoring bias has proven one of the most difficult biases to ameliorate (Furnham & Boo, 2011). The specific dual-process

characterization that will be used here contrasts "type 1" (t1) and "type 2" (t2) forms of cognition, because it is the most agnostic and thus inclusive extant dual-process characterization (Evans, 2008; Wason & Evans, 1975). Forms of cognition that are associated with conscious processes across multiple lines of literature can be generally identified as t2, and those associated with a lack of awareness can generally be considered t1.

The t1/t2 distinction arose within the reasoning literature, with the modern version largely based on the popular "System 1" (S1) vs. "System 2" (S2) distinction (e.g., Evans, 2003, 2008). In contrast to the S1/S2 distinction, t1/t2 suggests no assumptions regarding the evolutionary ordering and underlying neural correlates of the two modes of processing (Evans, 2008). The t1 versus t2 framework is also described here as "agnostic" because it accommodates different positions on other kinds of controversial issues from different dual-process literatures, and also their different terminologies and empirical foci. It allows for selectively ignoring details of the two types of cognition that are not immediately relevant to a given discussion, while including them in more general discussions. Shifting conjunctions among two constellations of characteristics jointly define these two general categories of cognitive process.

T2 processes occur serially, are capacity limited, and are dependent on attention, language, and/or working memory (Barrett, Tugade, & Engle, 2004; Frankish, 2009; Shiffrin & Schneider, 1977; cf., Shanks, 2003, 2005; Shanks, Rowland, & Ranger, 2005). T2 is tied to "higher-order" functions such as problem solving and reasoning, and refers to analytic, rulebased, systematic cognitive processes (e.g. Evans, 2003; 2008; Franksish & Evans, 2009; Kahneman & Frederick, 2002, 2006; Stanovich, Toplak, & West, 2008; Vygotsky, Hanfmann, & Vakar, 2012). Reflecting on the past and predicting behavioral outcomes are primarily t2controlled behaviors. Thus the use of rules, hypotheses, language, and logic are all t2-controlled behaviors.

T1 processes, on the other hand, have often been defined as any processes that fail to demonstrate t2 characteristics. They have thus been characterized by a lack of awareness, lack of an association with language, attention, or working memory, or their lack of capacity limitations. T1 processes include automatic, implicit, procedural, and associative aspects of cognition that take place in parallel and unintentionally. "Associative" refers both to the automatic aggregation of numerous simple associations, as well as the automatic activation of existing, well-established associations. Information that is learned in a t1 manner is less generalizable and less modifiable than t2-acquired information (Dienes & Berry, 1997; Lohse & Healy, 2012; Lohse & Ketels, 2012; Rickard, Healy, & Bourne, 1994; Roediger, 1990). T1 processes are unable to form new rules, hypotheses, or complex associations; that is, t1 forms only low-level associations between existing representations (Reder, Park, & Kieffaber, 2009).

Although t2 processes are needed to build novel complex associations, t1 continuously processes simple associations. T1 *processing* thus occurs very quickly. In behavioral output, this fast processing manifests as fast decision-making, described by some as relying on various heuristics. When output is primarily t2-controlled, the serial and deliberative nature of t2 slows behavioral output considerably as compared to cases of t1 control (Gilovich, Griffin, & Kahnemann, 2002; Kahnemann & Frederick, 2002). In contrast, in the case of *learning*, the fast processing of t1 manifests as slower learning. T1 continuously processes and aggregates any available contextual information in each new situation, which requires extremely fast updating (which requires fast processing). However, it is not directed like t2 processing, which instead focuses on whichever information is evaluated to be the most

relevant. Thus, the slower *updating* of t2, combined with its more efficient focus on only relevant information, results in much faster learning at the behavioral level: creating and applying categories, rules, and hypotheses leads to much more drastic behavioral change than incremental behavioral modification. T1 always learns indiscriminately, associatively, and statistically, slowly extracting environmental regularities from a vast input.

Independently of unit of analysis -- whether the individual, the task, or the interaction between the two -- there is an overall shift from t1 to t2, then back to t1, over the lifespan, within the context of a given task, and over the history of the species. This is not to say that complex behaviors are often "process pure," or subject to sole control by only one of the two forms of processing. In fact, complex real-world behaviors might *always* involve some degree of contribution from both t1 and t2 to behavioral control. In the course of learning a new task, we start out exploring the situation, reactively (t1). Then as we recognize stable action-outcome regularities we can begin to develop and test hypotheses (t2), to start behaving more proactively (t2) and eventually develop rules (t2) from hypotheses that prove predictive. After the repeated successful application of useful rules (t2), we can begin to automate the application of these rules (t1). As t1 processing leads to the identification of stable contextual regularities, control shifts to t2 for the processing of those regularities until conceptual representations deemed useful enough are automatically applied by t1. However, these stages occur in parallel - t1 and t2 processes behave competitively, cooperatively, or synergistically, but they do not wait in line to control behavior.

The lack of conscious accessibility makes the modification of t1 representations difficult. T2 representations can exclude contextual information that is not relevant, and can thus be applied across contexts. This independence leads to greater generalizability and directed

7

modifiability of representations that are built in this manner, as well as behaviors that are controlled in this manner (e.g., Squire & Wixted, 2011; Wohldmann & Healy, 2010; Wohldmann, Healy, & Bourne, 2008). However, again, learning involves the continuous interaction of t1 and t2 processes. When building new representations, the shifting relative contribution of t1 and t2 to behavior over the course of learning is an important titration that educators must negotiate, just as those who want to train humans to make decisions more rationally should design interventions that target behaviors that involve relatively more t1 or t2 control with process-appropriate training measures. For example, if the anchoring heuristic is applied in a primarily t1 manner, then affecting its use through t2 processes may require first making subjects aware of the details of anchoring, for manipulation by other t2 processes that lead to anchoring, then they should be more amenable to t2 debiasing approaches a priori.

The t1 vs. t2 distinction offers a useful semantic context for integrating discussions across multiple literatures. Here, it allows for the discussion of many disparate positions on dualprocess implications for anchoring, regardless of the specific theoretical dichotomy a given investigator focuses on (e.g., "intuitive" vs. "rational," "intuitive" vs. "conscious," "automatic" vs. "controlled," etc.). In the case of the heuristics and biases literature, rationality is associated with various other t2 processes and processing characteristics (although these vary among investigators). Thus the application of decisional heuristics that result in observable behavioral biases, such as the anchoring heuristic, is usually attributed to t1 processes and processing characteristics. In the case of the present experiments, a general prediction that arises from this perspective is that decisional biases operating in a t1 manner may not be amenable to directed modification by top-down, t2 control, in which case instructional debiasing would fail. The attribution of decisional biases to t1 processing is based in part in the difficulty in modifying them.

Anchoring, Primacy, & Recency

Anchoring effects in decision making and serial position effects in memory have both been studied extensively, albeit largely in isolation from one another (Atkinson & Shiffrin, 1968, 1971; Bonk & Healy, 2010; Cowan, Saults, Elliot, & Moreno, 2002, Davis, Rane, & Hiscock, 2013; Healy & Bonk, 2008; James, 1890; Nipher, 1878; Oberauer, 2003; Rundus, 1971; Waugh & Norman, 1965). "Serial position effects" in memory refer to better memory for items presented either earlier or later in lists than those in the middle. Primacy, or better memory for initial items than others, is the dominant finding, but recency, or better memory for the final item in each sequence, is also common. "Anchoring effects" refer to any decisional bias that arises from the anchoring heuristic. "Bias" here refers to a greater influence of particular items in a sequence than others, in cases in which all items should influence a subsequent decision equally for optimal performance. When subjects make decisions based on a list of sequentially presented items, primacy is more common that recency. Primacy in decision making is defined as a greater influence of the initial item or items on the subsequent decision. Hogarth and Einhorn's (1992) anchoring-and-adjustment model of belief updating also demonstrates how recency can obtain, given the presence of contextual conditions that facilitate its appearance. Primacy and recency in both recall and decision making have been described as arising from a combination of t1 and t2 processing.

Since early descriptions, primacy and recency effects are usually discussed as arising from separate cognitive mechanisms (e.g., Atkinson & Shiffrin, 1968, 1971; Rundus, 1971), some primarily t2- and some t1-controlled. Primacy in memory has often been conceptualized as

arising from t2-controlled aspects of behavior, such as rehearsal strategies involving starting at the beginning of the list of items (e.g., Atkinson & Shiffrin, 1968, 1971; Rundus, 1971). These strategies result in more time, and space in working memory, for the initially presented items than those presented subsequently (e.g., Atkinson & Shiffrin, 1968, 1971; Rundus, 1971). Observed primacy in memory has also been suggested to arise from constraints on the order in which items are required to be output, resulting in increased attention to, and processing of, earlier presented items (Cowan et al., 2002; Oberauer, 2003). If subjects are required to output items in the order of presentation, or if they do so despite no such requirement, then the structure of the task leads to more attention to the first item or items in a list. The classic account of recency in memory, on the other hand, suggests that the most recently presented item(s) are stored briefly in sensory memory, and thus the final one or two items are easily output immediately after list presentation, if there is no delay between presentation and recall (Atkinson & Shiffrin, 1968, 1971).

Anchoring, like other decisional biases, is often described as emerging from t1 processes, before information is processed declaratively (e.g., Chen & Bargh, 1997; Wilson et al., 1996). This suggestion is based on evidence of these biases even in decisions that are made too quickly for t2 control, as well as based on the subjects' inability to alter their application of these biases, even with incentives for better performance (e.g., Tversky & Kahneman, 1974; Wilson et al., 1996). However, if these biases are wholly t1, many researchers suggest that uninformative anchors should be as influential as informative anchors, but this is not always the case (Chapman & Johnson, 1999). Many descriptions of anchoring-and-adjustment describe the "adjustments" from the value of the initial representation, based wholly on the "anchor" value, as taking place in a t2 manner, with subjects explicitly adjusting their estimate based on each piece of new

information. However, these adjustments are overly conservative, so the final outcome cannot deviate far from the initial anchor. Recency obtains when subjects must update their Gestalt representation of their decisions after each new successive piece of information. One case of compulsory updating is based on the capacity limits of t2 processing: In the case of high stimulus complexity, information must repeatedly be consolidated to maintain an explicit representation of their decision based on the capacity limitations of t2. Thus increased stimulus complexity results in recency in decisions based on sequentially presented information (Hogarth & Einhorn, 1992).

The anchoring-and-adjustment model accounts for data from a wide range of contexts, and has been so dominant that investigators from many domains sometimes refer to the "anchoring" heuristic as the "anchoring-and-adjustment" heuristic (e.g., Arnott, 1998). However, some researchers have suggested limitations of the model. One example is Mussweiler and Strack (1999), who state that the anchoring-and-adjustment model requires the anchor to be outside the range of plausible values for the context of the decision. Stressing the importance of the semantic information offered in the specific context of the decision, these investigators have offered an alternative to the dominant anchoring-and-adjustment account, called the "Selective Accesibility Model" (e.g., Mussweiler, 2001; Mussweiler & Strack, 1999, 2001; Mussweiler, Strack, & Pfeiffer, 2000). Under this account, all information that is associated with a given anchor value is activated, and each successive piece of information reactivates that initial information, but only that which overlaps between the current item and the anchor. Thus it can be thought of as an activational confirmation bias – representations that are already active become more active. Under either the anchoring-and-adjustment or selective accessibility models,

confirmation bias can be considered as part of the cause of observed anchoring effects (e.g., Hogarth & Einhorn, 1992; Mussweiler & Strack 1999). "Confirmation bias" refers to the tendency to overtrust information that confirms previously held notions and undertrust disconfirmatory information (e.g., Klayman, 1995; Nickerson, 1998). In the case of decisions made with sequentially arriving information, the "previously held notion" is the influence of as much information as can be processed by t^2 – generally the first piece of information presented. However, anchoring and adjustment and selective accessibility make different predictions in regard to the differential importance of input order vs. output order. This is a question that has been addressed in the literature examining serial position effects in recall (e.g., Cowan, 2002; Oberauer. 2003), but not in the anchoring literature. The anchoring-and-adjustment account places special importance on the initially presented item, and so input order is implied to be the most important factor in determining which item is the anchor, and therefore whether primacy or recency obtains. The selective accessibility model, however, focuses on whichever information is most strongly activated, and thus if subjects focus primarily on an item that was presented later in the list, it should act as anchor, despite its not being the initial item in terms of presentation order.

Debiasing

Various attempts have been made to counteract decisional biases, including anchoring, but these attempts have often been unsuccessful, with a few exceptions (Arnott, 2006, Cook & Smallman, 2008, Furnham & Boo, 2011). Debiasing methods can target either the situation or the individual, and debiasing approaches that focus on the individual can use a relatively more t1- or t2-oriented approach. Debiasing manipulations in the present experiments involve attempting to influence individuals to cease reliance on anchoring in simple decisions based on sequentially presented information.

Decisional order effects are usually examined within the context of stimuli that carry semantic content, such as words and statements. Common cases in the literature include impressions of hypothetical people based on sets of sequentially presented adjectives, jury decisions based on sequentially arriving evidence, and judgments involving ropes and pulleys (Hogarth & Einhorn, 1992). Debiasing approaches often involve explicit instructions given to subjects, a t2 debiasing approach (Arnott, 2006; Cook & Smallman, 2008). Thus successful strategies are often specific to the particular experimental context of the decision. These include interventions such as having subjects consider multiple alternatives to the anchor, if alternatives are plausible with respect to the context of the decision (Hirt & Markman, 1995). Others have reported effective debiasing by having subjects consider the "opposite" of the anchor, which also varies across specific decision contexts (Chapman & Johnson, 1999; Galinsky & Mussweiler, 2001; Miller, Markman, Wagner, & Hunt, 2013; Mussweiler, Strack, & Pfeiffer, 2000). Other successful debiasing techniques have involved extensive description of the bias of interest, escalating in detail, first warning subjects about the bias, then describing the nature of the bias, followed by extensive training with feedback and continued declarative scaffolding (Arnott, 2006). In all successful cases though, debiasing interventions are specific to the specific context of the decision. If a general mechanism underlies anchoring, then a more domain-general debiasing method should exist.

However, on average, the results of even context-specific debiasing interventions have been inconsistent at best (Furnham & Boo, 2012). If decisional biases are based in t1 processing, this inconsistency may be due to the fact that many past debiasing efforts have either involved giving subjects explicit debiasing instructions, intended to modify aspects of behavior for which subjects lack t2 access (Cook & Smallman, 2008; Wilson et al., 1996), or altering these t1 biases without involving t2 processing at all (Arnott, 2006). On the other hand, if these biases are based in t2 processing, another explanation for debiasing failures is necessary. One possibility is that strategies such as instructional debiasing tend to put extra load on subjects' working memories; that is, they are told to "keep in mind" both the task at hand and the debiasing strategy that they are trying to implement (Cook & Smallman, 2008).

In the heuristics and biases literature, numerous general decisional biases have been instantiated across a number of specific contexts, and the same mechanisms are suggested to lead to these decisional biases across contexts and timescales (e.g., Chapman & Johnson, 1999). Thus strategies that prove successful in training away t1 decisional biases on a very short timescale in a simple laboratory task might also be effective for training away such biases on longer scales, and in complex, real world tasks, but only if interventions are applicable to a wide range of semantic contexts. Here we attempt both t2 and t1 approaches to debiasing anchoring, in simple centroid judgments, taking place on a very short timescale.

Chapter II Overview of Experiments

Five experiments were conducted within the context of an updated version of the centroid judgment task used in preliminary Experiments 1 & 2 (Appendix A, Wickens, Ketels, Healy, Buck-Gengler, & Bourne, 2010). In this task, subjects are asked to make centroid judgments, or to choose a single position on the screen that minimizes the average distance from their chosen location to each of the seven spatial locations presented on each trial. All five used our centroid judgment task, with 48 sequences of seven squares, displayed at seven different spatial locations within the bounds of a visible 20x20 grid on a computer screen (Figure 1). The first experiment focused on the effect of stimulus complexity on centroid judgments. The other four implemented different approaches to debiasing of the anchoring that is consistently evident in judgments in this paradigm.



Figure 1. The 20 x 20 grid as displayed to subjects, during feedback on firing and recall: Stimuli were presented one at a time on each trial, at seven different spatial locations within the grid. These stimuli were squares, displayed in one of four shades of red. The optimal deployment location was displayed as a green square during feedback, whereas subjects' chosen centroid was not displayed on the feedback screen. After choosing a centroid and recalling the seven items, this feedback screen was shown until subjects pressed a key to move on to the next trial.

Each new trial was automatically presented upon completion of the previous one. Each square was presented for 750 ms, with a square appearing at each location as the square at the previous location disappeared. During the first 24 trials, the training phase, feedback showing the ideal location was displayed immediately after each subjects' decision for each 7-item sequence. The centroid chosen by each subject was displayed as a turquoise square immediately after their decision, but was not displayed on the feedback screen. The ideal centroid was also displayed as a green square immediately after the centroid decision. Feedback on the decision was no longer given during the test phase (the second half of the trials). Short instructions appeared between training and test in all conditions, in all five experiments, warning subjects of this change. After each centroid judgment in Experiments 1-4, subjects were asked to recall, in any order, each of the seven locations, then the seven colors and sizes of each target. Feedback on item recall was presented in the four experiments that included the recall stage after the recall response for every item on a given trial, on every trial (both training and test), and in every condition. This feedback involved displaying the actual location, color, and size of each item, simultaneously, after all centroid and recall responses, until the subject pressed any key on the keyboard to advance to the next trial. This feedback screen also included the optimal deployment location, displayed as a green square (Figure 1).

The grid locations, colors, and sizes of the items displayed on each of the 48 trials were randomly generated, and then these seven specific stimuli were used across subjects and experiments. Thus on each trial, all subjects saw the same seven stimuli. Item presentation was counterbalanced in all five experiments in order to avoid any effects arising from the specific locations of each stimulus on a given trial. Specifically, for any subject, averaging across the 48 trials, the optimal deployment location could be closer to the items presented at a given serial position than to any other. In Experiments 2-5 we counterbalanced the first, second, third, fifth, sixth, and seventh items in each sequence across subjects using a balanced Latin square design, and thus completely avoided any such item effects. The counterbalancing scheme in Experiment 1 was slightly different than in the other four experiments because three conditions were included in the between-subjects comparison, and so this full counterbalancing would have resulted in too large a number of counterbalancing subgroups. Instead, the first and seventh, second and sixth, and third and fifth item locations were counterbalanced across subjects, for two counterbalancing subgroups within each of the three stimulus complexity conditions, and thus six groups in all.

Subjects were instructed to imagine themselves as Unmanned Aerial Vehicle (UAV) pilots in the military, that the onscreen grid was a representation of a battlefield, and that the stimuli were enemy locations. They were told their job was to place a UAV, only once, in a position to gather as much information as possible about as many enemies as possible. In the first four experiments, in which subjects were asked to recall stimulus characteristics, subjects were also told that after making their deployment decision they would have to relay information regarding the seven enemies from each sequence to other officers, and that this task was nearly as important as the deployment decision. There were four possible levels of stimulus color, ranging from grey to deeply red, and four levels of stimulus size. Size, color, and location were equally weighted in the calculation of the optimal deployment location (centroid judgment). The four possible enemy colors were described as representing the threat level of each enemy, whereas the four possible enemy sizes were described as representing the trustworthiness of the sources from which information regarding the enemies was gathered, and thus the certainty with which subjects could interpret information regarding that target. Subjects were told to place the

UAV in a position to collect as much information from as many enemies as possible, while collecting more information from enemies that represented a greater threat (more deeply red), and more information from those whose locations are more certain to be accurate (larger). Thus the cover story instructed subjects to make a centroid judgment in every case, but to weight their centroid judgments by the color and size of the targets when those stimulus characteristics varied.

Importantly, for each trial, the same specific seven item locations were used across the five experiments reported here. The specific colors and sizes of each stimulus presented at each spatial location were also the same for every subject, in those conditions in which size and color were varied. Thus the optimal deployment location was the same for every subject and for every trial, with the exception of the first two conditions in Experiment 1,¹ and conditions in Experiments 3 and 4 that involved the feedback debiasing manipulation. In all figures displaying bias in the decision (Figures 47, 10, 13, & 16), the distance of the optimal location from the item at each serial position is also plotted. Because Experiments 2-5 involved full counterbalancing across the six serial positions of interest, the distance of the optimal deployment location from all conditions without feedback debiasing is a straight line. In contrast, in Experiment 1, the optimal location deviates slightly from a straight line in all three conditions because the serial positions of three pairs of items were counterbalanced, rather than all six items of interest. Thus, counterbalanced pairs of serial positions (1/7, 2/6, & 3/5) are equidistant from the optimal deployment location, but across counterbalanced pairs there is some variance in distance from the optimal location. In conditions that included feedback debiasing during training, the distance

¹ In the first two conditions in Experiment 1, spatial location, and color and spatial location, respectively, were the only stimulus characteristics that varied and were thus the only stimulus characteristics used to calculate the ideal deployment location.

between the optimal deployment location and the location of the stimulus at each serial position also deviates from a straight line, because stimuli appearing at different serial positions were differentially weighted in calculating the centroid.

The task-relevant variation in color and size was added to increase the complexity of this simple decision, but was manipulated in the first experiment to examine any changes in decision behavior resulting from this additional stimulus complexity. Hogarth and Einhorn (1992) suggested an increase in the likelihood of recency with increased stimulus complexity, because high stimulus complexity results in updating representations to alleviate load on capacity-limited t2 processes. Although they described the range of stimulus complexity included in the investigations they reviewed, it remains unclear at which level of stimulus complexity recency begins to emerge.

We manipulated stimulus complexity in the manner described here in two preliminary experiments, described in Appendix A here, as well as in Wickens et al. (2010). This stimulus complexity manipulation was across a much smaller range of complexity than that reviewed by Hogarth and Einhorn (1992), but differences in stimulus complexity are more easily quantified in this paradigm than in previous tasks in which anchoring has been studied, most of which involve items with semantic content. Across these two experiments, we found that recency only emerged at the highest level of stimulus complexity, for which participants were asked to integrate item location, color, and size information in making decisions. Strong primacy was observed at all three levels of complexity, but *only* primacy was observed at the two lower levels of stimulus complexity. In the first experiment described here, we implemented this complexity manipulation in a single experiment, and in four others we implement various debiasing

manipulations, intended to decrease the influence of the initially presented items on each decision. These manipulations involved parts of the instructions, feedback debiasing, or both.

We examined two debiasing manipulations in Experiments 2, 3, and 4: instructional debiasing and feedback debiasing. In the feedback debiasing manipulation, feedback on the ideal centroid location was calculated with an algorithm that put much greater weight on whichever items were presented at the final serial positions in each sequence than the weight placed on those that occurred at the initial serial positions. Specifically, on each trial, the x and y coordinates of the optimal deployment location were calculated with the formulas in Figure 2.

$$\mathbf{x}_{o} = \frac{\sum_{i=1}^{7} (\mathbf{w}_{i}^{*} \mathbf{x}_{i} C_{i} S_{i})}{\sum_{i=1}^{7} (\mathbf{w}_{i}^{*} C_{i} S_{i})} \qquad \mathbf{y}_{o} = \frac{\sum_{i=1}^{7} (\mathbf{w}_{i}^{*} \mathbf{y}_{i} C_{i} S_{i})}{\sum_{i=1}^{7} (\mathbf{w}_{i}^{*} C_{i} S_{i})}$$

Figure 2. Calculation of the x and y coordinates of the optimal deployment location on each trial.

The calculation of the x and y coordinates for the optimal deployment location on each trial involved weighting the x and y coordinate of each stimulus (i) by the color (C) and size (S) of that stimulus, as well as a weighting factor for each serial position (w). Color and size each had four levels (1-4). The serial position weighting factor was 1/7 for each of the seven stimuli on a given trial in all conditions without feedback debiasing. However, in conditions of Experiments 3 and 4 that included feedback debiasing, *w* equaled 1.9/7 for the item at the seventh serial position, *w* equaled 1.5/7 for the sixth item presented, the fifth, fourth, and third items presented kept a weight of 1/7, the second item was given a weight of .5/7, and the first item presented was given a weight of .1/7. Thus the spatial position of the sixth item presented exerted 1/19 the influence of the spatial position of the seventh on the calculation of the ideal deployment location and the second item presented exerted 1/3 the influence of the sixth, and

 \sim 1/4 the influence of the last item. This information was only conveyed to subjects in the form of feedback on their decision, presented for 1000 ms, during the 24 training trials.

The instructional debiasing manipulation was an additional instruction, presented before the test phase in Experiments 2 and 3 but before both the training and test phases in Experiment 4, suggesting that subjects should keep in mind that information collected more temporally remotely is less reliable than information collected more recently, and thus they should factor information regarding more recently presented enemies into their UAV deployment to a greater extent than information presented earlier. In Experiment 5, we examined the effect of articulatory suppression on the centroid judgment, to test the hypothesis that anchoring in judgments based on sequentially arriving information arises from t1 processes (e.g., Chen & Bargh, 1997; Wilson et al., 1996). If this bias is indeed based in t1 processing, its form should be relatively unchanged by an articulatory suppression requirement, such as subjects repeating the word "Monday" aloud throughout the task. Table 1 shows the different specific conditions compared in each of the five experiments.

	Stimulus	D 11	Instructional	Instructional		A A
	characteristics	Recall	debiasing at	debiasing at	Feedback	Articulatory
	varied	required	training	test	deblashig	5uppression
E 1 - a	L	Х				
E 1 - b	LC	X				
E 1 - c	LCS	X				
E 2 - a	LCS	Х				
E 2 - b	LCS	Х		Х		
E 3 - a	LCS	X			Х	
E 3 – b	LCS	Х		X	Х	
E 4 – a	LCS	Х	X	X		
E 4 – b	LCS	Х	X	X	Х	
E 5 - a	LCS					
E 5 - b	LCS					Х

Table 1. Comparison of the conditions of the five experiments - A table tracking the manipulations in the five experiments reported here. L = variable location, LC = variable location & color, LCS = variable location, color, & size.

The five experiments described here examine decisional serial position effects within the context of a simple perceptual decision. This context allows an easily definable, objectively quantifiable, measure of the influence of each item on the decision. An advantage of this context is that it avoids the extra inferential step of rating items with subjective semantic content based on valence, as has been the case in a large number of investigations of anchoring involving decisions based on sequentially arriving semantic information (Hogarth & Einhorn, 1992). The dependent measure in all reported deployment bias analyses, for all five experiments, is the average city block distance, in pixels, of subjects' chosen centroid from the spatial position of the item presented at each serial position. Thus the data analyzed from each trial to examine the deployment bias are a time series of seven distances, with smaller distances between a chosen

centroid and an item at a given serial position interpreted as a greater influence of the information at that serial position on the centroid judgment. Counterbalancing involved the first, second, third, fifth, sixth, and seventh items, and so deployment bias and recall analyses included only these six items.

The within-subjects factors were also the same in all deployment bias analyses reported here. Primacy and recency were defined in terms of the interaction of two factors, "end" and "position". The factor of end had two levels, initial and final, comparing the average distance of the chosen centroid from the first three items relative to the average distance from the last three items. A simple effect of end suggests greater primacy relative to recency, or vice versa. This effect will be discussed as "simple primacy," or "simple recency." This definition is commensurate with discussions of these effects across various literatures that define "primacy" and "recency" only relative to one another (e.g., Hogarth & Einhorn, 1992) - under this definition, responses on a given trial can demonstrate either primacy or recency, but not both simultaneously. "Position" refers to the distance of a given item from the beginning or the end of the sequence in which it is presented, and thus had three levels, closest to the end (first and seventh items), second closest (second and sixth items), and third closest (third and fifth items). In this paradigm, the commonly found End x Position interaction usually results from a large influence of the initial item or two items on the centroid judgment - subjects choose centroids that are closer to the items nearer the ends of the sequence on average, but this tendency is more pronounced for the initial three than the final three items. When the data leading to an End x Position interaction follows this common pattern, it will be discussed as "complex primacy" (or "complex recency" if there are any cases in which the interaction is significant but the effect of end is driven primarily by the last three rather than the first three

items). We predicted that on average, subjects would remember better and deploy closer to items nearer the beginning and end of each sequence, and that this effect of position would be stronger for the initial three than the final three items in each sequence. This description reflects more of the detailed pattern of serial position effects that is usually observed in the memory literature on serial position effects (e.g., Atkinson & Shiffrin, 1968, 1971; Bonk & Healy, 2010; Healy & Bonk, 2008).

Across experiments, a main effect of end is thus indicative of a simple primacy effect, or chosen centroids that are closer, on average, to the first three than the last three presented items of each sequence, whereas an End x Position interaction demonstrates strong primacy, with chosen centroids closer to the first presented item than all others. Thus there were three within-subjects factors included in all deployment bias analyses: End, Position, and Phase (training and test). Recall analyses involved these same three factors. As mentioned earlier, the distance of the chosen centroid from the item at the fourth serial position was not included in any reported analyses, and only the between-subjects comparisons varied across the five experiments. In general, across experiments we predicted strong simple and complex primacy, with some cases of simple recency, in both deployment and recall. We also predicted various debiasing manipulations to alter these decisional serial position effects.

Chapter III Experiment 1

Experiment 1 examined the effect of stimulus complexity on decisional serial position effects. When decisions are made based on sequentially arriving stimuli, higher stimulus complexity is associated with a greater influence of more-recently presented items on the decision (Hogarth & Einhorn, 1992; Wickens et al., 2010). This result might arise from the added cognitive load of increased stimulus complexity, requiring updating a Gestalt representation of the information from a given trial after each new item is presented (e.g., Hogarth & Einhorn, 1992). This updating sets a new anchor each time, leading to an increased influence of the final item presented, as compared to situations in which this updating does not occur. Stimulus complexity was manipulated in the present experiment between-subjects, with three conditions. In the first condition, enemies varied only in their screen locations; in the second, they varied in both location and color; and in the third, they varied in location, color, and size. In the preliminary experiments described in Appendix A, we found primacy at all three levels of stimulus complexity, but recency only emerged at the highest level of stimulus complexity (Wickens et al., 2010). In those experiments however, although the two lower levels of complexity were included in a single experiment (preliminary Experiment 1), the highest level of stimulus complexity was only included in the second experiment, conducted a semester later than the first, and that was the only level of complexity included in preliminary Experiment 2. Experiment 1 was designed to replicate the effect of complexity found across those two experiments, in the context of an updated version of the computer program used for testing subjects. Appendix A describes the previous two experiments, whereas Appendix B includes a complete description of aspects of the present experiments that differ from those aspects from the previous set of experiments.

We predicted that subjects would show better recall for the first three items, on average, than the last three items presented in each sequence, and that those initial spatial locations would exert greater influence than the final spatial locations on centroid judgments, as demonstrated by a simple effect of end, or simple primacy, in the deployment bias analysis. These predictions are based on the literatures focusing on serial position effects in memory, in which primacy is a common finding, and anchoring effects in decision making, in which decisional primacy is a common finding. We also predicted complex primacy in both the recall and deployment bias analyses. That is, on average, we predicted that subjects would remember better and deploy closer to items nearer the beginning and end of each sequence, and that this effect of position would be stronger for the initial three than the final three items in each sequence. Results from previous versions of this paradigm support this prediction. We also predicted an increasing influence of the later items in the sequence with increasing stimulus complexity, as was the case in previous versions of the paradigm, including those reported in Appendix A. After withdrawing of feedback during the test phase, because there was no debiasing manipulation in this experiment, we predicted the pattern of bias to remain the same, with an overall performance increase in both recall and the decision from training to test, due to practice.

Method

Subjects

Sixty-six undergraduates from the University of Colorado at Boulder participated for course credit in the Fall 2012 semester. There were two counterbalancing subgroups within each of three levels of stimulus complexity, both manipulated between-subjects, for a total of six subgroups. Subjects were assigned by fixed rotation to one of those six subgroups, for 11 subjects per subgroup.

26

Design

Every subject saw the same randomly selected seven enemy positions on each trial. There were three complexity conditions: (a) *constant threat*, in which the threat level was the same for every target, (b) *variable threat*, in which the color, described as the level of threat each enemy represented, varied among targets on each trial, and (c) *variable threat and certainty*, in which the size of each target, described as representing the certainty of intelligence about that enemy, also varied among targets on each trial.

The first and seventh, second and sixth, and third and fifth item locations were counterbalanced across subjects, for two counterbalancing subgroups within each of the three stimulus complexity conditions, for six groups in all. However, counterbalancing subgroup is not included as a factor in any of the reported analyses. Thus deployment bias was analyzed with a 2x2x3x3 mixed factorial ANOVA, including within-subjects factors of end (initial and final), phase (training and test), and position (closest to end, second from end, third from end), as well as the between-subjects factor of stimulus complexity. Again, here in Experiment 1, the optimal deployment location deviates slightly from a straight line in all three conditions because the serial positions of three pairs of items were counterbalanced, rather than all six items of interest. The average distance of the items at each serial position from the optimal deployment location also varies somewhat among the three conditions in Experiment 1, because the calculation of the optimal deployment location was weighted by whichever stimulus characteristics were varied in each condition.

Analyses

Analyses are reported here for both the deployment and recall responses. The dependent measure in all three analyses was distance, measured in pixels, of subject responses to various

screen locations. For recall accuracy, the dependent measure was distance of each recall response from the enemy location that was closest to that response. Because subjects were not required to recall the items in the order that they appeared, it was impossible to know exactly which item a subject was attempting to recall with a given response. Thus, we assumed that the enemy location closest to a given recall response was the item that the subject was intending to recall with that recall response. In a limited number of cases, this resulted in a particular enemy location being compared to more than one recall location, for those cases in which two responses were closer to a particular enemy than any other. However, these cases were rare. For deployment accuracy, the dependent measure was the distance of the chosen deployment location from the single ideal deployment location on each trial. Finally, the influence of each item on the centroid judgment, or the biasing influence of the item presented at each serial position on "deployment," was defined as the distance from the chosen centroid to the spatial position of the item at each serial position. Again, the first and seventh, second and sixth, and third and fifth items were counterbalanced across subjects. This counterbalancing scheme allowed us to compare three pairs of items that were the same across subjects: The first (Serial Position 1), second (Serial Position 2), and third (Serial Position 3) items were compared to the last (Serial Position 7), second to last (Serial Position 6), and third to last (Serial Position 5) items, respectively. In this manner we compared the within-subjects variables of end (initial vs. final) and position (end, second to end, third to end). Thus any differences in the relative influence of the spatial positions of items appearing as either the initial three or final three serial positions is evidence of primacy or recency, respectively. The influence of the item at the fourth serial position was not included in any of the reported recall or deployment bias analyses, for any of the five experiments. However, the fourth item was included in the calculation of the optimal

deployment location on every trial, and in this way was factored into the deployment accuracy analyses.

Procedure

After completing a consent form, instructions were presented on the computer screen. Subjects then read the instructions, completed the 24 training trials, were presented the test phase instructions, completed the 24 test trials, and were then debriefed. Centroid and recall responses were made with the mouse, but the program only advanced past instructions pages, and between trials, with a key press on the keyboard. The battlefield was represented as a 20 x 20 black grid on a grey background. After the brief presentation of a green fixation cross in the center of the grid, seven square targets appeared sequentially, in pseudorandom locations on the screen, with each square representing an enemy location. Each item was presented for .75 s. Again, all seven spatial locations, stimulus colors, and stimulus sizes were randomly generated for each condition in which those stimulus dimensions varied, but these randomly selected stimulus dimensions were used for all subjects. Subjects were asked to make their deployment decision, and were then asked to recall the seven stimulus locations in any order. Finally, they were asked to recall the color and size of each enemy, in the order in which they recalled the spatial position of each enemy, by clicking on one of four response options in a box to the right of the battlefield grid. Thus on each trial, subjects (a) saw the seven stimuli at seven different spatial locations, (b) made their centroid decision, (c) received feedback on their decision, in the form of a green square marking the optimal deployment location, and a turquoise square marking their chosen centroid, (d) recalled the location of all seven items, (e) recalled the color and size of each of the seven items, and finally (f) received feedback on recall and deployment, in the form of a screen displaying the seven items and optimal deployment location
for that trial. Spatial recall responses were made by clicking on each recalled position within the bounds of the battlefield grid, whereas for color and size recall, a given recalled location was presented on the grid, and a prompt asked them to recall the color and then the size of each stimulus, by clicking in a box to the right of the grid displaying each of the four possible levels of color, then each of the four possible levels of size. After recall responses were collected for all seven locations, all seven colors, and all seven sizes, the actual seven stimuli from that trial were presented, simultaneously, until subjects pressed a key to continue. In addition, during the training trials, the ideal deployment location was also displayed during this feedback screen, again as a green square. Thus subjects were able to examine the relationship between the characteristics of the seven stimuli and the ideal deployment location, on each trial, for as long as they chose to do so. After the first 24 trials, subjects were warned that feedback on the decision would be withdrawn (although feedback on the color and size during the recall phase was maintained), and they then completed the second phase, the 24 test trials.

Results

Deployment Accuracy

As shown in Figure 3, in a 2 x 3 ANOVA with factors of phase and stimulus complexity, there were no significant effects (Fs < 1). Thus subjects did not improve significantly from training to test in their centroid judgments, despite a numerical improvement in all three complexity conditions. Although chosen centroids were numerically closer to optimal in the location, color, and size condition than in the location and color condition, and closer in both of these conditions than in the location-only condition, there were no significant effects involving stimulus complexity.

30



Figure 3. Deployment accuracy in Experiment 1. Error bars represent standard errors of the mean.

Deployment Bias

As can be seen in Figure 4, subjects deployed their UAVs closer to initial than final list items (F(1, 63) = 49.577, p < .0001, MSE = 836.591, $\eta^2 = .4404$), and nearer the items toward the ends of the lists (F(2, 126) = 44.063, p < .0001, MSE = 525.036, $\eta^2 = .4116$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2, 126)) = 28.358, p < .0001, MSE = 392.434, $\eta^2 = .3104$). There was also a significant effect of phase, because subjects deployed closer, on average, to the six critical locations during the test phase than during training (F(1, 63) = 33.756, p < .0001, MSE = 327.250, $\eta^2 = .3489$). There was also a significant Phase x End interaction because the difference between chosen centroids and initial and final items was larger in training than at test (F(1, 63) = 4.145, p = .0460, MSE = 440.878, η^2 = .0617). Finally, there was a significant Phase x Position interaction because subjects showed a less pronounced effect of position during test than during training F(2, 126) = 12.699, p < .0001, MSE = 121.222, $\eta^2 = .1678$). All these effects (with the exception of any effects involving the factor of phase) remained significant in an analysis including only the test phase (ps < .0001). There were no significant effects involving the stimulus complexity manipulation, despite a numeric trend during the test phase of greater recency with greater stimulus complexity (Figure 4). Also, the highest level of stimulus complexity, when location, threat (color), and certainty (size) were all varied (lcs), resulted in numerically less simple primacy (primacy relative to recency) than the other two levels of stimulus complexity.



Figure 4. Centroid judgment bias results from Experiment 1.

Recall

As can be seen in Figure 5, subjects recalled initially presented stimulus locations better than final (F(1, 63) = 31.693, p < .0001, MSE = 84.849, $\eta^2 = .3447$) and remembered items toward the ends of the lists better than those nearer the middle (F(2, 126) = 14.402, p < .0001, MSE = 42.324, $\eta^2 = .1861$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2, 126) = 4.274, p = .0160, MSE = 49.553, $\eta^2 = .0635$). There was a significant effect of phase (F(1, 63) = 4.426, p = .0394, MSE = 243.489, $\eta^2 = .0656$) because subjects recalled the six items of interest slightly better at test than during training, and

also a significant Phase x End interaction because subjects showed less-pronounced simple primacy during test than during training (F(1, 63) = 18.922, p < .0001, MSE = 48.884, $\eta^2 =$.2310), with the location only condition even demonstrating numerical recency during test. Of greatest interest were any differences among the three levels of stimulus complexity. The only such effect was a significant End x Complexity interaction (F(2,63) = 4.671, p = .0128, MSE =84.849, $\eta^2 = .0690$). The variable threat (location and color), and variable threat and certainty (location, color, and size) conditions showed attenuated primacy at test, with the constant threat condition even showing numerically more recency than primacy. In an analysis of only the test phase, the only effect that remained significant was the effect of position, because on average subjects recalled items toward the ends better than those toward the middle of sequences $(F(2,126) = 7.950, p = .0006, MSE = 32.850, \eta^2 = .1121)$. The effects of end (F(1,63) = 2.843, p = .0006, MSE = .0006,.0967, MSE = 80.855, $\eta^2 = .0432$) and End x Complexity (F(2,63) = 2.399, p = .099, MSE =80.855, $\eta^2 = .0708$) remained only marginally significant in this analysis of the test phase on its own. This interaction reflects the difference between the strong simple primacy, on average, in the variable threat condition, weaker simple primacy in the variable threat and variable certainty condition, and lack of simple primacy, on average, in the constant threat condition.



Figure 5. Recall performance of subjects in Experiment 2.

Discussion

In Figure 4, more simple recency is evident in the condition of highest stimulus complexity, in both training and test, than in the other two conditions. However, this trend is only numeric, because there were no significant effects involving the stimulus complexity manipulation for deployment bias. Figure 5 shows that the same was not true for recall. In that case, the lcs condition showed slight simple primacy, the variable threat condition showed more simple primacy, and the constant threat condition showed slight simple recency with the pattern of complex primacy for the first three items presented. This effect is thus taken as simple recency, with an effect of the factor of position. At test, subjects in the lc condition recalled items that were displayed at the fifth, sixth, and seventh serial positions better on average than the first three, but they recalled the first item as well as the final three. Most definitions of order effects

in decision making literatures would call this pattern of results neither primacy nor recency. The stronger primacy in the constant threat condition led to a significant end x stimulus complexity interaction in the recall analysis, but this was the only hint of evidence for Hogarth and Einhorn's (1992) suggestion, and the detailed pattern of results was not what they would have predicted. The deployment accuracy analysis demonstrated no significant effect of either factor, nor their interaction (phase and complexity), but the trend was mostly sensible, with better deployment accuracy during test than during training, and subjects deploying closer, on average, to the optimal centroid in the condition of highest stimulus complexity than the other two. The lack of effects involving complexity here may be due to a lack of adequate statistical power, but all effect sizes were very small. Thus there is unlikely to be an effect to detect, even a very small one. It is likely that the range of stimulus complexity tested here was all within what Hogarth and Einhorn (1992) would have classified as "low complexity," and thus did not lead to sufficient load to force representation updating. The pattern of results reported in Appendix A thus likely arose from the specific items used in the high complexity condition, as they were different than in the two lower complexity conditions in that case. However, because anchoring occurred at all three levels of stimulus complexity, Experiments 2-5 included only the highest level of stimulus complexity, with variation in enemy location, color, and size. This level of complexity requires subjects to integrate the three streams of information in making their centroid judgments, and is thus more similar to real world decisions than variation in only location or in location and color.

Chapter IV Experiment 2

Experiment 2 examined the effect of instructional (t2) debiasing on the anchoring evident in this decision. Two debiasing groups were included. One was not subject to a debiasing manipulation, whereas the other was given debiasing instructions before the 24 test trials. The "no debiasing" condition was identical to the high complexity condition in Experiment 1, and thus was shown the same test phase instructions as in Experiment 1, only warning subjects that deployment feedback would cease to be displayed. Before the test phase, the debiasing group was shown instructions suggesting greater weight for each successive item in a given sequence. Specifically, the debiasing instructions stated: "You are engaged in integrating intelligence information from sources collected over time. As is typical in the dynamic hostile environment, things change, and so earlier arriving information is less reliable than information which has just been received." Thus something like a linear weighting was described, within the context of the cover story, suggesting placing a successively greater weight on each item in a given sequence. However, in both conditions, deployment feedback on each trial during training, which again marked the optimal deployment location with a green square, reflected the same equal influence of each of the seven items that was used in Experiment 1.

We expected debiasing instructions between training and test to result in attenuated primacy and more recency in chosen centroids. We also expected training performance in both conditions to be similar, because the manipulation only took place after the training phase. Finally, we expected subjects to improve in deployment and recall accuracy from training to test, and we expected patterns of primacy in recall accuracy and deployment bias that are similar to those found in Experiment 1.

Method

Subjects

One hundred and eight undergraduates from the University of Colorado Boulder participated for course credit in this experiment in the Spring 2013 semester. They were assigned by a fixed rotation to one of the 12 counterbalancing subgroups, with nine subjects in each subgroup, and thus 54 subjects in each condition.

Design and Procedure

The design and procedure of Experiment 2 were the same as those of Experiment 1, with two important exceptions. First, as already described, the between-subjects manipulation compared instructional debiasing before the test phase to no such debiasing manipulation. All subjects were in a condition equivalent to the highest level of stimulus complexity from Experiment 1, under which subjects are required to integrate information regarding the color, size, and screen location of each target in making their centroid judgments.

The second change was that we counterbalanced the first, second, third, fifth, sixth, and seventh items in each sequence across subjects using a balanced Latin square design. Thus the optimal centroid on each trial was equidistant from each of the locations at the six serial positions included in the analyses, when averaging across the six counterbalancing subgroups. Because there were three levels of the between-subjects manipulation in Experiment 1, it was not practical to use this more complete counterbalancing scheme in that experiment. However, Experiments 2-5 each only included a single two-level between-subjects manipulation; thus this more complete Latin-square counterbalancing of the six items of interest was included in all subsequent experiments.

Results

Deployment Accuracy

As shown in Figure 6, in a 2 x 3 ANOVA with factors of phase (training and test) and debiasing (debiasing instructions and no debiasing instructions), there were no significant effects involving debiasing (Fs < 1). However, subjects' chosen centroids were closer to optimal, on average, during the test phase than during training (F(1, 106) = 7.580, p = .0069, MSE = 201.771, $\eta^2 = .0667$). Although this difference was numerically greater in the instructional debiasing condition than in the no debiasing condition, the Phase x Debiasing interaction was not significant (F < 1).



Figure 6. Deployment accuracy in Experiment 2. Error bars represent standard errors of the mean.

Deployment Bias

As can be seen in Figure 7, subjects deployed their UAVs closer to initial than to final list items (F(1, 106) = 59.124, p < .0001, MSE = 715.779, $\eta^2 = .3581$), and nearer the items toward the ends of the lists (F(2, 212) = 51.581, p < .0001, MSE = 632.183, $\eta^2 = .3273$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2, 212) = 51.581).

212) = 41.105, p < .0001, MSE = 464.318, n² = .279). There was also a significant effect of phase, because subjects deployed closer, on average, to the six critical items during the test phase than during training ($F(1, 106) = 135.302, p < .0001, MSE = 268.237, n^2 = .5607$). Finally, there was a significant Phase x End x Position interaction because subjects showed lesspronounced primacy during test than during training (F(2, 212) = 9.173, p = .0002, MSE =167.217, $\eta^2 = .0796$). All these effects (with the exception of any effects involving the factor of phase) remained significant in an analysis including only the test phase (ps < .0001). In the cases of both the general analysis and the analysis including only the test phase, there were no significant effects involving the between-subjects debiasing manipulation (Fs < 2, ps > .2). However, in the first, more inclusive analysis including the factor of phase, the End x Debiasing interaction was marginally significant (F(1, 106) = 3.034, p = .0844, MSE = 715.779, $\eta^2 =$.0278), because the simple primacy effect (indicated by the effect of end) was attenuated in the instructional debiasing condition as compared to the no debiasing condition. However, this difference between conditions was no longer significant in the analysis of only the test phase (p >.2).



Figure 7. Centroid judgment bias results from Experiment 2.

Recall

As can be seen in Figure 8, subjects recalled initially presented stimulus locations better than final ($F(1, 106) = 99.242, p < .0001, MSE = 84.503, \eta^2 = .4835$), and remembered items toward the ends of the lists better than those nearer the middle (F(2, 212) = 26.169, p < .0001, $MSE = 34.956, \eta^2 = .1979$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction ($F(2, 212) = 9.514, p = .0001, MSE = 40.857, \eta^2 = .0824$). There were also significant interactions of Phase x End (F(1, 106) = 13.601, p = .0004, MSE = $49.627, \eta^2 = .1137$) and Phase x End x Position ($F(2, 212) = 9.554, p = .0001, MSE = 27.885, \eta^2 = .0827$), because subjects showed less-pronounced primacy, both simple and complex, during test than during training. In an analysis including only the test trials, there was also a marginally significant Position x Debiasing interaction ($F(2, 212) = 2.799, MSE = 28.004, p = .0631, \eta^2 =$.0257). The effect of position was attenuated in the instructional debiasing condition, because in this condition subjects recalled the second and third items from each end nearly as well as the first.



Figure 8. Recall performance of subjects in Experiment 2.

Discussion

Subjects in Experiment 2 demonstrated the predicted improvement from training to test in their centroid judgments, apparent in Figure 6. However, the effect of phase was also present in the deployment bias analysis. Thus, both of these effects are likely due to item effects, because the optimal centroids in the 24 test trials were closer, on average, to the seven item locations than centroids were during the training trials. This difference results in less room for any reasonable attempt at selecting a centroid to vary, because it is the result of the seven specific items on each trial being closer to each other, on average. Thus deployment accuracy was likely also impacted

by this effect of the specific items used. However, it is relatively unsurprising in an easy task that most learning occurs in early trials. All of the bias effects seen in Experiment 1 were present in both conditions, but the effect of end was attenuated in the instructional debiasing condition to some extent, although this attenuation was only marginally significant. However, differences during training, before the instructional debiasing manipulation took place, contributed to this marginally significant attenuation. Thus the only clear interpretation is that the instructional debiasing manipulation did not work in this case. Given the difficulty of ameliorating the influence of anchoring in other decisions, this outcome should not be overly surprising. However, if adjustments from an initial anchor are made in a t2 manner, as many investigators have suggested, then an instructional debiasing intervention should have some effect on the outcome of the decision. This was not the case. If anchoring arises from a combination of t1 and t2 processes (e.g., Epley & Gilovich, 2005), then a debiasing intervention involving both t1 and t2 components might be necessary.

Chapter V Experiment 3

Experiment 3 compared debiasing through weighting the feedback on deployment decisions during training ("feedback debiasing" group), a primarily t1 intervention, to a debiasing intervention that combines feedback debiasing with the instructional debiasing from Experiment 2 ("feedback and instructional debiasing" group), a combined t1 and t2 debiasing intervention. The feedback debiasing manipulation is classified here as t1 because subjects had to learn the specific serial position weighting scheme, and this learning likely took place in the absence of awareness (t1). Subjects were instructed to explicitly attend to the spatial relationship between the centroid location that they chose, the optimal deployment location, the seven targets and their characteristics, but they were not given the actual weights assigned to each serial position in calculating the optimal deployment location. The "feedback debiasing" group was shown the same test phase instructions as the "no debiasing" group from Experiment 2 (and all three conditions from Experiment 1), only warning that feedback on deployment would cease. The feedback and instructional debiasing group was shown the test phase instructions from the "debiasing" group in Experiment 2, suggesting greater weight for each successive item in a given sequence. We predicted feedback debiasing during training to result in attenuated primacy and more recency in chosen centroids than in Experiments 2 and 3. However, we predicted that debiasing instructions before test for half the subjects would result in even more attenuation of primacy in the deployment, because this condition combined t1 and t2 approaches to debiasing. We also predicted training performance in both conditions to be similar, because the instructional debiasing manipulation only took place after the training phase. Finally, we predicted that subjects would improve in deployment and recall accuracy from training to test,

and we predicted patterns of primacy in recall accuracy and deployment bias to be similar to those found in Experiments 1 and 2.

Method

Subjects

Seventy-two undergraduates from the University of Colorado Boulder participated for course credit in this experiment in the Spring 2013 semester. They were assigned by a fixed rotation to one of the 12 counterbalancing subgroups, with six subjects in each subgroup, 36 subjects in each debiasing condition.

Design and Procedure

The design and procedure of Experiment 3 were exactly the same as those of Experiment 2, with the addition of weighted feedback on the deployment decision on each trial during training. Again, instead of a serial position weighting factor of 1/7 for each of the seven serial positions, the item at the seventh serial position was given a weighting factor of 1.9/7, the sixth item presented was given a weighting factor of 1.5/7, the fifth, fourth, and third items presented kept a weight of 1/7, the second item was given a weight of .5/7, and the first item presented was given a weight of .1/7.

Results

Deployment Accuracy

As shown in Figure 9, in a 2 x 3 ANOVA with factors of phase (training and test) and debiasing (feedback debiasing and feedback debiasing during training with debiasing instructions at test), subjects' chosen centroids were closer to optimal, on average, during the test phase than during training (F(1, 70) = 22.696, p < .0001, MSE = 173.488, $\eta^2 = .2448$). This difference was greater in the feedback and instructional debiasing condition than the feedback debiasing only

condition, resulting in a significant Phase x Debiasing interaction (F(1, 70) = 5.486, p = .0220, $MSE = 173.488, \eta^2 = .0727$). Importantly, this interaction is not a result of item effects, because the items were identical in the two conditions, but it is driven by a pre-existing, chance difference between the two groups before any manipulation took place.



Figure 9. Deployment accuracy in Experiment 3. Error bars represent standard errors of the mean.

Deployment Bias

As can be seen in Figure 10, subjects deployed their UAVs closer to initial than final list items (F(1, 70) = 63.625, p < .0001, MSE = 652.813, $\eta^2 = .4761$), and nearer the items toward the ends of the lists (F(2, 140) = 39.859, p < .0001, MSE = 679.940, $\eta^2 = .3628$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2,140) = 29.141, p < .0001, MSE = 554.472, $\eta^2 = .2939$). There was also a significant main effect of phase, because subjects deployed closer, on average, to the six critical items during the test phase than during training (F(1, 70) = 101.733, p < .0001, MSE = 237.989, $\eta^2 = .5924$). There was a significant Phase x End x Position interaction because subjects showed less-pronounced complex primacy during test than during training (F(2, 140) = 7.910, p = .0006, MSE = 266.645, $\eta^2 = .1015$). There were also significant End x Debiasing (F(1, 70) = 4.456, p = .0384, MSE = 652.813, $\eta^2 = .0598$) and Position x Debiasing (F(2, 140) = 4.778, p = .0098, MSE = 679.940, $\eta^2 = .0639$) interactions, because, contrary to predictions, the feedback and instructional debiasing group demonstrated a *more* pronounced primacy bias than the feedback debiasing group, during both test and training. In another repeated measures ANOVA including only data from the test phase, again the Position x Debiasing interaction was marginally significant (F(2, 140) = 2.932, p = .0566, MSE = 440.558, $\eta^2 = .0402$), but not any other effects involving debiasing. In this analysis, all other reported effects in the analysis including both phases remained significant (ps < .0001), with the exception, of course, of the reported effects involving the factor of phase.

In two ANOVAs analyzing the two conditions separately, all the reported effects from the general analysis remained significant in both conditions; however, in the feedback and instructional debiasing condition, two marginally significant effects emerged, a Phase x Position interaction (F(2, 70) = 2.601, MSE = 354.346, p = .0814, $\eta^2 = .0692$), and a Phase x End interaction (F(1, 35) = 3.548, MSE = 548.433, p = .0679, $\eta^2 = .0920$). All of these effects can be seen in Figure 10. Although not included in any analyses, in Figure 10, data from the "no debiasing" group in Experiment 2 is compared to the two groups in this experiment.



Figure 10. Centroid judgment bias results from Experiment 3, as compared to the "no debiasing" condition from Experiment 2.

Recall

As can be seen in Figure 11, subjects recalled initially presented stimulus locations better than final (F(1, 70) = 33.112, p < .0001, MSE = 93.082, $\eta^2 = .3211$), and recalled items nearer the ends of the lists better than those toward the middle (F(2, 140) = 13.012, p < .0001, MSE =47.853, $\eta^2 = .1567$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2, 140) = 6.593, p = .0018, MSE = 47.348, $\eta^2 = .0861$). There was also a significant Phase x End interaction because subjects showed less-pronounced primacy, both simple and complex, during test than during training (F(1, 70) = 5.862, p = .0181, MSE =39.308, $\eta^2 = .0773$). Unlike the other experiments reported here, there was also a significant interaction of Phase x Position x Debiasing (F(2, 140) = 4.516, p = .0126, MSE = 26.788, $\eta^2 =$.0606), because the effect of position was attenuated more from training to test in the feedback debiasing group than in the feedback and instructional debiasing group. There was also a significant Phase x End x Position x Debiasing interaction (F(2, 140) = 3.192, p = .0441, MSE =30.256, $\eta^2 = .0436$), because the greater position effect attenuation from training to test in the feedback debiasing group was only evident for the initial three items. There were no other interactions involving debiasing group.



Figure 11. Recall performance of subjects in Experiment 4, as compared to the "no debiasing" condition from Experiment 2.

Discussion

The results of Experiment 3 are more difficult to interpret than the results of the first two experiments, largely because of pre-existing differences between the two groups of subjects. The feedback and instructional debiasing group performed more poorly on the decision than the other group during training, when there was no manipulation differentiating the two conditions. This is likely the primary reason for the significant Phase x Debiasing interaction in the deployment accuracy analysis – that group improved more from training to test because they had more room to improve. Based on a qualitative comparison of the two conditions without instructional debiasing from Experiments 2 and 3, the feedback debiasing manipulation appears to have no effect. It is possible that subjects completely ignored the feedback on their deployment decision. In Experiment 4, we added debiasing instructions before training as well, intended to (a) draw subjects' attention to the debiasing feedback to a greater extent than in other iterations of the paradigm, and (b) strengthen any effect of instructional debiasing.

Chapter VI Experiment 4

Experiment 4 involved a comparison of a condition involving instructional debiasing to a condition involving both feedback and instructional debiasing. Thus it was a comparison of the second condition from Experiment 2 (Instructional debiasing alone) to the second condition from Experiment 3 (Feedback and instructional debiasing), within the context of one experiment. Experiment 3 involves the first comparison of a condition with feedback debiasing vs. a condition without, and therefore involves the comparison of performance in two conditions with different optimal deployment locations, one based on a weighting scheme placing greater importance on items occurring later in each sequence, another with the normal equal weighting of the item characteristics for each of the seven items. Another difference between the first three experiments and Experiment 4 is that the instructional debiasing in this experiment was included in both the initial instructions and then again before test, rather than only before test as was the case in all conditions involving instructional debiasing reported thus far. The specific wording of the instructions was changed slightly, to fit the syntax of the initial instructions, and to fit within the new context in the case of the debiasing instructions before training (because the debiasing instructions before test were, in this case, a reminder rather than an introduction to the idea of weighting later items to a greater extent in centroid calculations). The extra debiasing instructions, presented before training, were intended to strengthen the effect of both debiasing manipulations. Presenting debiasing instructions before both training and test should strengthen t2, instructional debiasing simply because the debiasing instructions were given twice. However, the extra instructional debiasing was also expected to strengthen the effect of feedback debiasing, by drawing subjects attention to it, synergistically resulting in more effective debiasing than in Experiment 3. If subjects are told to weight later occurring items more than initial, then they

might be able to interpret feedback that is weighted in a similar manner more effectively. This procedure is akin to successful past debiasing approaches involving extensive declaratively-scaffolded training with immediate feedback on each decision (Arnott, 2006).

Method

Subjects

Eighty-four undergraduates from the University of Colorado Boulder participated for course credit in this experiment in the Spring 2013 semester. They were assigned by a fixed rotation to one of the 12 counterbalancing subgroups, with 7 subjects in each subgroup, 42 subjects in each condition.

Results

Deployment Accuracy

As shown in Figure 12, in a 2 x 3 ANOVA with factors of phase (training vs. test) and debiasing (instructional debiasing at training and test vs. feedback debiasing with instructional debiasing at training and test), subjects' chosen centroids were closer to optimal, on average, during the test phase than during training (F(1, 82) = 4.493, p = .0371, MSE = 256.420, $\eta^2 = .0603$). There were no significant effects involving the debiasing manipulation (ps > .10). However, the decrease from training to test in the distance of subjects' chosen deployment location from optimal was numerically more pronounced in the condition that included feedback debiasing during training.



Figure 12. Deployment accuracy in Experiment 4. Error bars represent standard errors of the mean.

Deployment Bias

As can be seen in Figure 13, subjects deployed their UAVs closer to initial than final list items ($F(1, 82) = 31.388, p < .0001, MSE = 554.244, \eta^2 = .2768$), and nearer the items toward the ends of the lists ($F(2, 164) = 48.697, p < .0001, MSE = 452.328, \eta^2 = .3726$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction (F(2, $164) = 24.185, p < .0001, MSE = 422.427, \eta^2 = .2278$). There was also a significant effect of phase, because subjects deployed closer, on average, to the six critical items during the test phase than during training ($F(1, 82) = 75.735, p < .0001, MSE = 334.297, \eta^2 = .4801$). There was also a significant Phase x End interaction because there was a greater difference between chosen centroids and initial and final items during training than during test (F(1, 82) = 11.622, p =.0010, $MSE = 354.497, \eta^2 = .1241$). There was a significant Phase x Position interaction because on average subjects showed a less pronounced effect of position during test than during training ($F(2, 164) = 4.938, p = .0083, MSE = 296.809, \eta^2 = .0568$). Finally, there was a significant Phase x End x Position interaction because subjects showed less-pronounced complex primacy during test than during training (*F*(2, 164) = 12.549, *p* < .0001, *MSE* = 182.136, η^2 = .1327). There were no significant effects involving debiasing group (*Fs*<1).

In an analysis involving only the test phase, all the reported effects in the analysis including both phases remained significant, with the exception, of course, of those effects involving the factor of phase (ps < .05). However, as shown in Figure 13, data from the "no debiasing" group in Experiment 2 demonstrated simple primacy that was stronger than in the two conditions in Experiment 4. A marginally significant End x Experiment interaction in an analysis including the first 42 of those Experiment 2 subjects, offers some support for this effect (F(2,129) = 2.892, p = .0591, MSE = 714.556, $\eta^2 = .0429$).



Figure 13. Centroid judgment bias results from Experiment 4, as compared to the "no debiasing" condition from Experiment 2.

Recall

As can be seen in Figure 14, subjects recalled initially presented stimulus locations better than final ($F(1, 82) = 50.322, p < .0001, MSE = 95.039, \eta^2 = .3803$), and recalled items nearer ends of the lists better than those toward the middle (F(2, 164) = 33.173, p < .0001, MSE = $36.426, \eta^2 = .2880$). This effect of position was greater for the initial items, as evidenced by an End x Position interaction ($F(2, 164) = 6.459, p = .0020, MSE = 48.342, \eta^2 = .0730$). There were also significant interactions of Phase x End ($F(1, 82) = 21.488, p < .0001, MSE = 51.941, \eta^2 =$.2076), Phase x Position ($F(2, 164) = 5.292, p = .0059, MSE = 24.744, \eta^2 = .0606$), and Phase x End x Position ($F(2, 164) = 8.956, p = .0002, MSE = 27.371, \eta^2 = .0985$), because subjects showed less-pronounced primacy, both simple and complex, during test than during training. There were no significant effects involving debiasing (Fs < 2, ps > .2).



Figure 14. Recall performance of subjects in Experiment 4, as compared to the "no debiasing" condition from Experiment 2.

Discussion

In Experiment 4, the general pattern of results observed in the first three experiments was evident, including the lack of any effect of debiasing manipulations. However, comparing the two conditions from Experiment 4 to the no debiasing condition from Experiment 2, we found some evidence of an effect of debiasing instructions presented before both training and test. This marginal debiasing success may have occurred because the debiasing manipulations in Experiment 4 were closest to the optimal titration of t1 and t2 aspects of debiasing that has proved successful in the past (e.g., Arnott, 2006), and closer to the manner in which t1 and t2 contribute to behavioral control in complex naturalistic situations.

However, the debiasing manipulations from Experiments 2, 3, and 4 were all designed based on the assumption that establishing an anchor is a t1 process, with each subsequent adjustment taking place in a t2 manner. Under the selective accessibility model (e.g., Mussweiler & Strack, 1999), the relative contributions of t1 and t2 are very similar – hypothesis testing takes place with relatively more t2 control, whereas the activation of anchor-related representations occurs automatically, or under t1 control. The feedback debiasing manipulation was intended to ameliorate anchoring at the same preverbal stage that gives rise to anchoring. The instructional debiasing manipulations were intended to alter their explicit learning strategy, thus changing the adjustment or hypothesis testing behaviors that lead to anchoring in the decision. Experiment 5 was intended to help illuminate the confused state of description across various literatures regarding the different roles of t1 and t2 in leading to anchoring in decisions.

Chapter VII Experiment 5

Experiment 5 was much like Experiment 2, in that a condition with no debiasing manipulation was compared to a condition intended to alter subjects' decisional bias. However, there were two important differences. First, unlike the other four experiments, there was no recall requirement in this experiment. We dropped the recall requirement to increase subject engagement in the task, simplify the design, and thus focus on bias in the decision in the absence of extra requirements that could alter performance. Second, we did not include feedback or instructional debiasing in any form. Thus subjects in both conditions in this experiment were given deployment feedback on each of the 24 training trials based on equal weight ascribed to the characteristics of enemies occurring at each of the seven serial positions. Third, the debiasing manipulation involved articulatory suppression rather than a debiasing training strategy. If anchoring in these decisions is a t1 decisional bias, as it has been described, then this manipulation should not affect anchoring. Under almost every description of the effect of articulatory suppression offered by previous investigators, occupying articulatory processes in this manner loads some t2 process (e.g., declarative processing, working memory, language, etc.), forcing subjects to complete tasks in some t1 manner (e.g., procedurally, in the absence of working memory, implicitly, etc.

Method

Subjects

Forty-eight undergraduates from the University of Colorado Boulder participated for course credit in this experiment in the Spring 2013 semester. They were assigned by a fixed rotation to one of the 12 counterbalancing subgroups, with 4 subjects in each subgroup, 24 subjects in each condition.

Articulatory Suppression

Subjects were required to repeat the word "Monday," out loud, from the point at which they finished reading the initial instructions through the end of all 48 trials. They were allowed to repeat the word as quickly or as slowly as they chose, the only requirement was that they created an audible, continuous stream of speech throughout the 48 trials. To ensure that participants met these requirements, an experimenter sat nearby and listened. In the condition without articulatory suppression, the experimenter sat in the same place in case the presence of the experimenter affected performance in any way. The experimenter corrected subjects when they either: a. spoke too quietly to be heard, or b. paused for too long between recited "Monday"s.

Results

Deployment Accuracy

As shown in Figure 15, in a 2 x 3 ANOVA with factors of phase and stimulus complexity, there were no significant effects (Fs < 1). Thus subjects did not improve significantly from training to test in their centroid judgments, and articulatory suppression did not change deployment accuracy.



Figure 15. Deployment accuracy in Experiment 5.

Deployment Bias

As can be seen in Figure 16, subjects deployed their UAVs closer to initial than final list items (F(1, 46) = 11.185, p = .0016, MSE = 59.503, $\eta^2 = .1956$). There was also a significant effect of phase, because subjects deployed closer, on average, to the six critical items during the test phase than during training (F(1, 46) = 412.201, p < .0001, MSE = 61.165, $\eta^2 = .9493$). There was also a marginally significant End x Debiasing interaction (F(1, 46) = 3.082, p = .0858, MSE = 59.503, $\eta^2 = .0628$). As was the case in the previous experiments, subjects in the no debiasing condition deployed nearer the first three than the last three items, whereas subjects in the articulatory suppression condition did not. This was the only effect involving debiasing that approached significance in this experiment, and there were no significant effects involving position, because subjects did not deploy nearer the items closer to the ends of each sequence (Fs < 1).

By far the largest stable effect was the difference between training and test, but this was the result of item effects – the optimal deployment location was closer to the average position of

the six items of interest in the test phase than during training. However, separate analyses were conducted on training alone, test alone, as well as on each of two debiasing conditions in the absence of the other. In the analysis involving only the test phase from both groups of subjects, the simple effect of end remained marginally significant (F(1,46) = 4.372, p = .0842, MSE = 120.590, $\eta^2 = .0868$), although the End x Debiasing interaction did not (p > .1). In the analysis of only the training trials, there were no significant effects (Fs < 2, ps > .15). Thus in training, although subjects chose centroids that were numerically closer to the beginning three than the final three items, this difference was not significant. Nonetheless, the addition of these observations from training in the analysis including both training and test allowed the marginally significant effect of end during the test trials to reach significance in the overall analysis.



Figure 16. Centroid judgment bias results from Experiment 5, as compared to the "no debiasing" condition from Experiment 2.

However, in Figure 16, it appears, qualitatively, that this effect of end is only present in the no debiasing condition, without any appearance of an end effect in the articulatory suppression condition. Results of analyses on the two conditions separately, each including both training and test trials, confirmed that this trend was significant. The simple effects of phase $(F(1,23) = 167.570, p < .0001, MSE = 76.536, \eta^2 = .8793)$, and end $(F(1,23) = 10.106, p = .0042, MSE = 76.571, \eta^2 = .3053)$ both remained significant in the no debiasing condition, but only the effect of phase remained significant in the articulatory suppression condition $(F(1,23) = 270.538, p < .0001, MSE = 45.794, \eta^2 = .9216)$. With articulatory suppression, the effect of end was not significant $(F(1,23) = 1.770, p = .1964, MSE = 42.435, \eta^2 = .0715)$.



Figure 17. Deployment accuracy comparison across conditions from Experiments 1, 2, and 5. All three conditions included variation in stimulus size, color, and position, and no debiasing manipulation. Although the lack of a recall component resulted in better centroid judgment accuracy in Experiment 5, than the other two experiments, all three conditions show a qualitatively similar pattern, with a small improvement from training to test.

Discussion

Based on subject reports, the task was more engaging without item recall. The recall component also added a working memory requirement that was not present in Experiment 5, and

this led to markedly better centroid judgment accuracy and less demonstrated decisional bias in this than in any other experiment (Figures 14 & 15). However, with this much smaller variation in end and position effects, articulatory suppression resulted in no primacy whatsoever. Possible explanations for this difference are discussed below. The effect of articulatory suppression is the most striking finding from any of the five experiments because it goes against suggestions that anchoring in decision making is a primarily t1 decisional bias (e.g., Chen & Bargh, 1997; Wilson et al., 1996). If anchoring is caused by primarily t1 processes, then it should not have been affected by a declarative dual-task manipulation. However, articulatory suppression might only have interfered with the t2-controlled "adjustment" or "hypothesis-testing" phase of decision making in this task, and thus ameliorated the effect of anchoring without actually preventing the anchor from being established in the first place, instead only impeding the overt behavioral strategies that lead to the inordinate influence of the anchor. Thus these results suggest a significant role of t2 processes in anchoring, although the stage of processing at which they exert their effects is unclear. However, these t2 processes are independent of the specific semantic context of the decision, because this effect was found in the context of this simple perceptual decision.

Chapter VIII Conditions Without Debiasing

Conditions in three of the five experiments involved no debiasing manipulations. These three conditions were compared in two repeated measures ANOVAs examining deployment bias and deployment accuracy. The variable threat and variable certainty (location, color, and size) from Experiment 1, and the two "no debiasing" conditions from Experiments 2 and 5, were included in these analyses. The effects of interest in these analyses involve the factor of experiment. This accuracy analysis was intended to verify that in the absence of recall, subjects in Experiment 5 chose centroids closer to optimal than subjects in the other experiments, as was the case in a previous experiment without item recall, the pattern of decisional primacy was not significantly different in the comparison between conditions with and without recall. Nonetheless, the deployment bias analysis was included to examine any differences in primacy in deployment between Experiment 5 and the other four experiments. Only effects involving the factor of experiment will be reported here.

Deployment Accuracy

Two versions of the accuracy analysis were included, one with 22 subjects from Experiment 1, all 54 from Experiment 2, and all 24 from Experiment 5, and another including the same 22 subjects from Experiment 1, the first 24 subjects from Experiment 2, and the same 24 subjects from Experiment 5 (Figure 17). Results were almost identical in the two analyses, so only the analysis with nearly equal N is discussed here. The effect of experiment was the only significant difference (F(2, 67) = 13.198, MSE = 1087.801, p < .0001, $\eta^2 = 2.826$), verifying that subjects chose centroids that were closer, on average, to the ideal location in Experiment 5 than in Experiment 1 and Experiment 2. Deployment accuracy was almost identical in Experiment 1 and Experiment 2.

Deployment Bias

The same 70 subjects were included in the deployment bias analysis who were included in the accuracy analysis. Deployment bias performance of this subset of subjects from Experiment 2 was similar to deployment bias performance in Experiment 1, so all the differences reported here were a result of the comparison between Experiment 5 and the first two. There was a significant effect of experiment, because subjects deployed closer, on average, to the six counterbalanced items in Experiment 5 than in Experiment 1 and Experiment 2 (F(2,67) = 9.499, MSE = 1457.039, p = .0002, $\eta^2 = .2209$). There were also significant interactions of End x Experiment (F(2,67) = 4.208, p = .0190, MSE = 417.225, $\eta^2 = .1115$), Position x Experiment $(F(4, 134) = 5.523, p = .0004, MSE = 363.577, \eta^2 = .1415)$, and End x Position x Experiment $(F(4,134) = 3.863, MSE = 367.507, p = .0053, n^2 = .1034)$, because subjects in Experiment 5 exhibited attenuated simple primacy as compared to subjects in Experiment 1 and Experiment 2, did not demonstrate the effect of position that was evident in deployment in the first two experiments, and thus did not exhibit the complex primacy that was evident in the first two experiments. Finally, the interaction of Phase x End x Position x Experiment was marginally significant (F(4,134) = 2.324, p = .0598, MSE = 126.875, $\eta^2 = .0649$), because the pattern of bias changed very little from training to test in Experiment 5, whereas complex primacy was stronger in training than during test in the first two experiments.

Chaper IX General Discussion

Although some marginally significant differences in deployment behavior resulted from the feedback and instructional debiasing approaches that we examined, articulatory suppression and the inclusion of the recall component were by far the strongest two manipulations. This may be a result of a fundamental change in the manner in which subjects completed the task - a change forced by occupying declarative processing and/or working memory. Although many decisional biases may arise from t1 processing, these results suggest that anchoring in decisions made with sequentially arriving information arises from a t2 processing approach in this task. Because the task itself is a simple perceptual decision without semantic content, the t2 components of anchoring in this task are not limited to the specific semantic context of the decision. Instead, there is fundamental role of t2 processing in decisions involving the integration of sequentially arriving information, even in decisions that do not necessarily involve semantic content.

In Experiment 5, the articulatory suppression manipulation involved repeating the word "Monday," continuously throughout the 48 trials. This form of articulatory suppression likely selectively loads some, but not all, t2 processes. Specifically, articulatory suppression occupies declarative processing while attention is left unstrained. In contrast, the recall requirement in Experiments 1-4 loads working memory, but not necessarily declarative processes. In the absence of this recall requirement, in Experiment 5, subjects perform much better on the deployment decision overall. Experiments 2-4 also collectively suggest some degree of efficacy of a simple instructional debiasing manipulation, whereas the feedback debiasing manipulation appears to be completely unsuccessful.
Effects established using previous versions of the paradigm involving the factors of end and phase were replicated in the first four experiments. These effects are pure serial position effects, independent of any influence of specific item characteristics, because item position was counterbalanced across subjects (fully counterbalanced in Experiments 2-5). Thus subjects in the first four experiments deployed closer to and remembered items better that were nearer to the ends of each sequence than to any others. However, in the deployment, this effect was usually driven by the distance from chosen centroids to the first item presented on each trial. Deployment locations were much closer to the first item on average than any others, and subjects remembered that item better than any others. They also remembered the first three items presented in each sequence better than the last three items, and deployed closer to those initial items. This effect of end in all experiments was not wholly driven by the average distance of the chosen centroid from the first item. In contrast with the other four experiments, position and the interaction of End x Position were not significant in Experiment 5, although we again found an effect of end.

Experiment 1 failed to demonstrate the effect of stimulus complexity on the decision that was evident in preliminary Experiment 1, despite a numeric trend in the predicted direction (Figure 4). There are at least three possible reasons for the failure to replicate this effect. First, accuracy performance on the deployment decision varied more in Experiment 1 than in Experiments 2-4, despite the similarity between Experiment 1 and each of these other experiments. This variability might have been due in part to differences in the subjects used – Experiment 1 used subjects at the end of a semester, soon before the final deadline for subjects to fulfill their credit for general psychology. These subjects may have been less engaged in the task than subjects used during the following semester, well before this deadline. The pattern of

results shown in Figure 17 qualitatively supports this suggestion. Comparing identical conditions from Experiments 1 and 2, both with no debiasing manipulation, the mean deployment accuracies from Experiment 1 show more variance than the same condition in Experiment 2, based on the error bars displaying standard errors of the means (SEMs). Second, with only 11 subjects per counterbalancing subgroup and 22 subjects per condition, the smaller n per condition in this experiment may also be partially to blame. However, the error in Experiment 2 was much more similar to that in Experiments 3 and 4, both of which had closer to the number of subjects per conditions as the 22 per condition in Experiment 1, with standard errors in the first experiment almost twice as large as in the other two no debiasing conditions, from Experiments 2 and 5 (Figure 17). Thus a combination of these explanations may be best. Given the amount of error, the number of subjects in Experiment 1 may have been inadequate to detect the effect of complexity in the decision. Finally, the effect of complexity that we established across the preliminary experiments may have arisen from the specific items used in the condition of highest stimulus complexity, because the two preliminary experiments used completely different sets of stimuli. Thus the optimal deployment location may simply have been closer to the final item, on average, in the lcs condition than in the other two. In any case, the effect of complexity is unnecessary to establish effects of the debiasing manipulations examined in Experiments 2-5.

Subjects in Experiment 5 deployed closer, on average, to the ideal location than did subjects in the other four experiments. This may be in part due to the attenuated primacy bias in this situation - subjects deploy closer, on average, to each of the seven items, and thus deploy closer to the single ideal location. Thus they did not demonstrate the more complex pattern of primacy that obtains in this paradigm when the recall requirement is included. Instead, subjects

67

in this experiment only demonstrated simple primacy, with deployment locations closer to the initial three than final three items presented in each sequence. Subjects in Experiment 5 demonstrated no effect of position on the decision, and no interactions involving this factor. Importantly though, there was an interaction of End x Debiasing that was marginally significant, and analyses of the two conditions independently demonstrated a significant effect of end in only the condition without articulatory suppression. Thus although the recall requirement itself greatly reduced accuracy and made complex primacy emerge, simple primacy was only absent in the articulatory suppression condition, across the five experiments.

Many investigators have suggested that output processes are at least partly responsible for observed serial position effects in memory (e.g., Cowan et al., 2002; Oberauer, 2003). The recall component itself may be to blame for the strong serial position effects that are evident in most versions of this paradigm. There were no constraints on the order of item recall; subjects were told to recall the items "in any order," but many subjects output the first presented item location first. However, if the recall requirement was solely responsible for serial position effects in the decision, because the items that were recalled best exerted the largest influence on the decision, then the pattern of results in the identical analyses of recall and deployment bias should be the same. This is not the case.

Another explanation is that the recall requirement loads working memory in a way that the decision without recall does not, because subjects must "keep in mind" the specific item locations and other stimulus characteristics. Without this requirement, in Experiment 5, subjects reported relying on a more Gestalt strategy, involving "making their eyes fuzzy" in order to "get the gist" of the seven items on a given trial to choose the weighted centroid. These subject selfreports thus suggest decision differences resulting from the recall requirement that cannot be explained solely by output requirements. Empirical precedent in multiple lines of literature supports this suggestion, with evidence that bias in decisions based on sequentially arriving information is independent of memory for the specific items that contribute to the decision (e.g., Hastie & Park, 1986; Waldron, Patrick, Duggan, Banbury, & Howes, 2008). In fact, the approach to the deployment decision that subjects reported in Experiment 5 was very similar to a strategy that was explicitly taught to pilots in an experiment by Waldron and colleagues (2008), called "triangulation". That experiment showed that this triangulation strategy resulted in worse memory for specific items, because subjects abstracted the centroid without specifically attending to specific item locations.

The Effect of Practice

Across the first four experiments reported here, subjects demonstrated numerically less bias in both decision and recall during test than during training, with significant interactions involving the factor of phase in some cases. These effects are commensurate with past work on decisional biases in experts making decisions within their domain of expertise. Although since the first description of the anchoring heuristic investigators have noted that even experts, within their domain of expertise, are susceptible to these biases (Tversky & Kahnemann, 1974), these biases are less severe in experts than in novices (Furnham & Boo, 2011).

In the context of a simple centroid judgment task, "expertise" is easily achieved. Thus, even within the short span of 24 training trials, subjects reach asymptotic performance very quickly. As subjects gain expertise, they are likely able to process the information on each trial with increasing efficiency, thus decreasing the overall cognitive load created by each trial. This decrease in cognitive load allows subjects to take all seven targets into account increasingly evenly, because attention is not over-allotted to initially presented items.

Theoretical Implications

Two accounts of serial position effects in decisions based on sequentially arriving information have emerged as the most influential, the anchoring-and-adjustment model (e.g., Hogarth & Einhorn, 1992; Tversky & Kahnemann, 1974) and the selective accessibility model (e.g., Mussweiler & Strack, 1999). Although the two are not necessarily mutually exclusive, they do make somewhat different predictions in the current context. First, within terms from the literature involving serial position effects in memory, because the selective accessibility model stresses the importance of activation, and output order would be predicted to be more important than input order, whereas the importance of the absolute position of the anchor – that it is the initially presented item – is stressed in the anchoring-and-adjustment model, and so input order is implied to be most important. If input order is the basis of anchoring however, large differences in performance between conditions requiring recall versus those that did not require recall should not obtain. Nevertheless, this comparison demonstrated larger differences than any other, across the five experiments. Under both accounts, articulatory suppression would be predicted to ameliorate anchoring, because both adjustment and hypothesis testing processes are described by many as occurring in a t2 manner (e.g., Furnham & Boo, 2011). Thus the marginal effect of this manipulation further distinguishes the dominant descriptions of these two models from less popular accounts describing both stages of anchoring – both setting the anchor and adjusting from that point - as t1.

Hogarth and Einhorn (1992) taxonomized decisions involving sequentially arriving information based on whether a decision involves estimation or evaluation, the level of stimulus complexity, and whether decisions are step-by-step or end-of-sequence. "Estimation" refers to decisions that involve the averaging of sequentially presented stimuli, whereas "evaluation" decisions refer to decisions in which subjects must integrate sequentially arriving pieces of information into a single decision in regard to a particular hypothesis for which that information is relevant (e.g., guilty vs. innocent in jury decisions). Thus the question of whether primacy or recency will obtain in a given decision depends on whether subjects must update their decision representation repeatedly or only at the end of item presentation. Hogarth and Einhorn suggested that updating a single representation of all the information in a given trial after each new piece of evidence leads to averaging of the stimulus values, resulting in recency. In that case each successive item exerts an increasing influence on the final impression, and thus recency occurs. Simple, one-shot decisions involve accumulating all the pieces of evidence on a given trial and then making one decision at the end. This approach leads to primacy, because capacity limitations in attention and working memory limit the amount of information that can be stored. Under this account, the strong primacy effect in the decision that we find in the first four experiments occurs because this end-of-sequence estimation decision involved simple stimuli. All three of these task characteristics are associated with primacy, but not recency, arising from anchoring (Hogarth & Einhorn, 1992), because none of the three characteristics force subjects to update representations after each successive item. Under this explanation, the lack of recall in Experiment 5 relieved WM load compared to the other experiments, and thus although the decision was still an end-of-sequence estimation decision based on simple stimuli, it was so easy that only simple primacy remained.

The selective accessibility model accounts for the present results slightly better than does the anchoring-and-adjustment model. In this task, overlapping features between stimuli are easily quantified, given that there are only three features along which the stimuli vary (location, color, and size). Thus every anchor activates representations associated with one of the four

71

levels of stimulus color, one of the four levels of stimulus size, and some amount of spatial information within the space of the battlefield grid. This automatic activation could lead to a chosen centroid that is closer to the initial item presented in each sequence, on average, than any other, even without any contribution from t2 processing. The drastic difference in deployment decision between the no debiasing condition in Experiment 5 as compared to similar conditions in Experiments 1 and 2 supports this account. The recall requirement leads to a t2 recall strategy, which itself leads to greater activation for the characteristics of the first item presented. Thus when recall is included, strong complex primacy emerges that is not evident in the absence of any recall requirement. This result is commensurate with findings that suggest that an increase in subject engagement and effort can, under certain conditions, actually increase the severity of anchoring in simple decisions (e.g., Epley & Gilovich, 2005).

The most likely explanation of the present results involves a combination of t1 and t2 processes leading to anchoring across the many tasks that have been used to study anchoring, with their relative contributions varying based on the specific experimental context. The centrality of semantic information in making the specific decision examined is likely the most important moderator of the strength of serial position effects observed in decisions based on sequentially arriving information. However, even in a task whose core does not fundamentally involve semantic information and without a recall requirement, a t2 approach leads to some degree of anchoring, manifest as a simple effect of end in deployment bias. The marginal effect of articulatory suppression in ameliorating this simple effect supports this conclusion.

As is the case of memory for lists of sequentially presented semantic stimuli (e.g., Atkinson & Shiffrin, 1968, 1971; Rundus, 1971), primacy arises from t2 processes in the sense that some form of explicit rehearsal occurs. In the present task however, it is visospatial rehearsal – picturing the items where they just appeared. Although this form of rehearsal does not involve language, it involves a controlled, deliberate, strategic approach to learning, and thus can be called primarily t2. For most individuals, this rehearsal occurs in the order of item presentation. Thus the first item is remembered better than any other item, this memory interferes with memory for the other items, and the first item exerts a larger influence on the decision than does any other item. The decision without recall is controlled with a greater t1 contribution to behavior because it allows for a Gestalt approach to the deployment decision by not focusing subjects' attention on each individual item, and thus complex primacy is eliminated. Articulatory suppression in Experiment 5 occupies t2 processes, shifting primary control to t1, and thus eliminates simple primacy as well. Under this explanation, order effects arise from t2 processing approaches, not t1 processes. Other investigators have framed anchoring within the context of dual-process theory, basing their theory on the suggestions of Hogarth and Einhorn (1992). They suggest that explicit information integration processes result in stimulus averaging, and primacy, in information integration (e.g., Betsch, Kaufmann, Lindow, Plessner, & Hoffmann, 2006).

Conclusion

The basic pattern of results here is robust; there is a strong one-item anchoring effect that was evident in each of the four experiments requiring recall. Thus these results may generalize to other situations – at least part of anchoring is based on general characteristics of human decision making and the limited capacity human cognitive system. In general, more contact should be made between the anchoring literature and various educational literatures. Investigators in both domains might find useful insights from the other. Appendix C includes brief speculation on educational applications of the present results and implications from the anchoring literature more generally. However, in the present experiments there were small or nonexistent effects of various debiasing manipulations. Thus more work is needed to find the appropriate balance of t1 and t2 in developing a general debiasing strategy that can ameliorate anchoring across tasks, but a general technique that reduces anchoring across contexts may be an unrealistic goal.

References

- Anderson, N. H. (1981). Foundations of Information Integration Theory. New York, Academic Press.
- Arnott, D. (1998). A taxonomy of decision biases. *Technical Report, Monash University, School of Information Management and Systems, Caulfield.*
- Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal*, 16, 55-78. doi: <u>10.1111/j.1365-2575.2006.00208.x</u>
- Aron, L. Y. (2006). An overview of alternative education. Urban Institute, Washington, D.C.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation*, *2*, 89-195.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control processes of short-term memory. Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.
- Baars, B.J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.
- Ball, D. L., Sleep, L., Boerst, T. A., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *The Elementary School Journal*, 109, 458-474.
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130, 553-73. doi:10.1037/0033-2909.130.4.553
- Betsch, T., Kaufmann, M., Lindow, F., Plessner, H., & Hoffmann, K. (2006). Different principles of information integration in implicit and explicit attitude formation. *European journal of social psychology*, *36*, 887-905.
- Block, J. H., (Ed.). (1971). *Mastery learning, theory and practice*. New York: Holt, Rinehart, & Winston.
- Block, J. H. & Burns, R. B. (1976). Mastery Learning. Review of Research in Education, 4, 3-49.
- Bonk, W. J., & Healy, A. F. (2010). Learning and memory for sequences of pictures, words, and spatial locations: An exploration of serial position effects. *American Journal of Psychology*, 123, 137-168.

- Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*, 98-113.
- Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational behavior and human decision processes*, 79, 115-153.
- Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The selffulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, 33(5), 541-560.
- Clement, J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions Finding anchoring conceptions for grounding instruction on student's intuitions. *International Journal of Science Education*, 11, 554-565. 10.1080/0950069890110507
- Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors*, 10, 745-754.
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, 46, 153-177. <u>http://dx.doi.org/10.1006/jmla.2001.2805</u>.
- Davis, L. C., Rane, S., & Hiscock, M. (2013). Serial recall of visuospatial and verbal information with and without material-specific interference: Implications for contemporary models of working memory. *Memory*, (ahead-of-print), 1-20.
- Davis, D., & Sorrell, J. (1995). Mastery learning in public schools. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved [December, 1995] from http://teach.valdosta.edu/whuitt/files/mastlear.html.
- Dewey, J. (2007). Experience and education. Free Press.
- Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*, 3-23. doi:10.3758/BF03210769
- Epley, N. & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, 18, 199–212.
- Evans, J. S. B.T. (2003). In two minds: dual-process accounts of reasoning *Trends in Cognitive Sciences*, 7, 454-459.
- Evans, J. S. B.T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Reviews of Psychology*, 59, 255-278.

- Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal distinction. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds: Dual Process and Beyond* (pp. 89-108). New York, NY, US: Oxford University Press, Inc.
- Frankish, K., & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In J. S. B. T. Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond (pp. 1–30). Oxford, UK: Oxford University Press.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40, 35-42.
- Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: the role of perspective-taking and negotiator focus. *Journal of personality and social psychology*, *81*(4), 657.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology* of intuitive judgment. New York: Cambridge University Press.
- Gurpinar, E., Musal, B., Aksakoglu, G., & Ucku, R. (2005). Comparison of knowledge scores of medical students in problem-based learning and traditional curriculum on public health topics. *BMC Medical Education*, 5, 7.
- Hammer, D. (1996). Misconceptions or P-Prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *The Journal of the Learning Sciences, 5,* 92-127.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 258-268.
- Healy, A. F., & Bonk, W. J. (2008). Serial learning. In H. L. Roediger, III (Ed.), Cognitive psychology of memory (pp. 53-63), Vol. 2 of Learning and memory: A Comprehensive reference, 4 vols. (J. Byrne, Editor). Oxford: Elsevier.
- Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*, 1069.
- Hogarth, R., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1-55. doi:10.1016/0010-0285(92)90002-J
- Holland, S., Griffiths, R., & Woodman, M. (1997, March). Avoiding object misconceptions. In *ACM SIGCSE Bulletin*, 29, 1, 131-134. ACM.
- James, W. (1890). Principles of psychology. New York: Holt.

- Jaynes, J. (1976) *The origin of consciousness in the breakdown of the bicameral mind*. Houghton Mifflin.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Ed.s) Heuristics of Intuitive Judgment: Extensions and Applications, Cambridge University Press, Cambridge, GB.
- Kahneman, D., & Frederick, S. (2006). Frames and brains: Elicitation and control of response tendencies. *Trends in Cognitive Sciences*, 11, 1-2.
- Ketels, S. L., Healy, A. F., Wickens, C. D., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2011). *A dual-process account of decision making: Memory and anchoring.* Manuscript submitted for publication.
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, *32*, 385-418.
- LeBoeuf, R. A., & Shafir, E. (2009). Anchoring on the" here" and" now" in time and distance judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 81.
- Lohse, K. R., & Healy, A. F. (2012). Exploring the contributions of declarative and procedural information to training: A test of the procedural reinstatement principle. *Journal of Applied Research in Memory and Cognition*, *1*, 65-72. doi:10.1016/j.jarmac.2012.02.002
- Lohse, K. R. & Ketels, S. L. (2012). Implications of dual-process theories for optimizing motor learning and performance. In A. L. Magnusson & D. J. Lindberg (Eds). *Psychology of performance and defeat* (pp. 65-90), Hauppauge, NY: Nova Science Publishers.
- Margetson, D. B. (1999). The relation between understanding and practice in problem-based medical education. *MEDICAL EDUCATION-OXFORD*, 33, 359-364.
- Miller, A. K., Markman, K. D., Wagner, M. M., & Hunt, A. N. (2013). Mental simulation and sexual prejudice reduction: the debiasing role of counterfactual thinking. *Journal of Applied Social Psychology*, 43, 190-194.
- Mussweiler, T., & Strack, F. (1999). Comparing is believing: A selective accessibility model of judgmental anchoring. *European review of social psychology*, *10*, 135-167.
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, 26(9), 1142-1150.
- Mussweiler, T. (2001). The durability of anchoring effects. *European Journal of Social Psychology*, *31*(4), 431-442.

- Mussweiler, T., & Strack, F. (2001). The semantics of anchoring. *Organizational Behavior and Human Decision Processes*, *86*(2), 234-255.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review* of General Psychology, 2, 175.
- Nipher, F. E. (1878). On the distribution of errors in numbers written from memory. Transactions of the Academy of Science of St. Louis, 3, CCX–CCXI.
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49, 469-483.
- Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy* (1st ed.). London: Routledge.
- Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness : Reinterpreting memory in terms of activation and binding. *Psychological Bulletin*, 135, 23-49. doi:10.1037/a0013974
- Rickard, T. C., Healy, A. F., & Bourne Jr., L. E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 1139–1153.

Roediger, H. L. (1990). Retention without remembering. American Psychologist, 45, 1043-1056.

- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63-77.
- Samuels, R. (2009). The magical number two, plus or minus: Dual process theory as a theory of cognitive kinds. In K. Frankish & J. S. B. T. Evans (Eds.), *In two minds: dual process and beyond* (pp. 1-27). Oxford, UK: Oxford University Press.
- Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience*, 4, 244-256. doi:10.1162/jocn.1992.4.3.244
- Shanks, D. R. (2003). Attention and awareness in "implicit" sequence learning. In L. Jimenez (Ed.), *Attention and implicit learning* (pp. 11–42). John Benjamins Publishing Company.
- Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (202-220). London: Sage.

- Shanks, D. R., Rowland, L. A., & Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological research*, 69, 369-82. doi:10.1007/s00426-004-0211-8
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127-190.
- Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*. 119, 3–2226.
- Smith, A., Lovatt, M., & Turner, J. (2009). *Learning to Learn In Practice:* The L2 Approach. Crown House Publishing Bethel, CT.
- Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *The Annual Review of Neuroscience*, *34*, 259-288.
- Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior*, 251-285.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
- Vygotsky, L., Hanfmann, E., & Vakar, G. (2012). *Thought and language*. Cambridge, MA: M.I.T. Press.
- Waldron, S. M., Patrick, J., Duggan, G. B., Banbury, S., & Howes, A. (2008). Designing information fusion for the encoding of visual–spatial information. *Ergonomics*, 51, 775-797.
- Wason, P. C., St, J., & Evans, T. (1975). Dual processes in reasoning? Cognition, 3, 141-154.
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. Psychological review, 72, 89.
- Wickens, C. D., Ketels, S. L., Healy, A. F., Buck-Gengler, C. J., & Bourne, L. E. (2010). The anchoring heuristic in intelligence integration: A bias in need of de-biasing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54, 2324-2328.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology-General*, 125, 387-402.
- Wohldmann, E. L., & Healy, A. F. (2010). Exploring specificity of speeded aiming movements: Examining different measures of transfer. *Memory & Cognition*, 38, 344-55. doi:10.3758/MC.38.3.344

Wohldmann, E. L., Healy, A. F., & Bourne, L. E. (2008). A mental practice superiority effect: less retroactive interference and more transfer than physical practice. *Journal of experimental psychology: Learning, memory, and cognition*, 34, 823-33. doi:10.1037/0278-7393.34.4.823

Appendix A - Preliminary Experiments

This appendix describes preliminary Experiments 1 and 2. Many, but not all, of the results reported here were published in Wickens et al. (2010). In preliminary Experiment 1, we examined integration complexity and transfer in decisions involving sequentially presented information. In preliminary Experiment 2, we added another level of complexity, and examined the effect of an instructional debiasing manipulation on recall and the decision.

Preliminary Experiment 1

Preliminary Experiment 1 examined learning and retention in a spatial list learning and decision paradigm, as well as a t1 decisional bias that operates on the decision. Specifically, we used a paradigm in which each trial consisted of seven sequentially presented spatial locations followed by a decision requiring information integration. Previous work with this paradigm demonstrated a strong influence of primacy, or anchoring, on the decision (e.g., Ketels et al., 2011). Subjects imagined themselves as army intelligence officers who had to place a small unmanned aerial vehicle (UAV) on a battlefield to gather information. Squares marked the seven spatial locations, representing enemies on the battlefield, a 20x20 grid on a computer screen. We increased stimulus complexity, as compared to previous versions of the paradigm (Ketels et al., 2011), in the form of four possible threat levels for each enemy. The enemies could appear in one of four colors, varying from grey to red, representing these four levels of threat. On 48 trials in each of two sessions, spaced 1 week apart, subjects were told to place the UAV in a position to collect as much information from as many enemies as possible, while collecting more information from enemies that represented a greater threat. Subjects were then asked to reproduce the location and threat level of each enemy.

Based on our previous results (Ketels et al., 2011), we hypothesized that for the deployment decision, the initial spatial locations would exert greater influence than the final spatial locations. We expected this bias to decrease from the beginning of the first week to the beginning of the next, as subjects gained expertise. Again, we also expected recency to gradually emerge across the two levels of stimulus complexity, with decisions in the location-only condition demonstrating less bias than those in the location and color condition.

Method

Subjects

Thirty-six undergraduates from the University of Colorado at Boulder participated for course credit in the Spring 2009 semester. They were assigned by a fixed rotation to the two recall conditions, and to one of six counterbalancing subgroups in each condition. Thus there were 12 subjects in each recall condition and 2 subjects in each subgroup.

Design

Every subject saw the same randomly selected seven enemy positions on each trial. The spatial positions of the first, second, third, fifth, sixth, and seventh enemies were counterbalanced by a balanced Latin Square across subjects. There were two complexity conditions: (a) *variable threat*, in which the color, called "threat level," varied among targets on each trial, and (b) *constant threat*, in which the threat level was the same for every target. The specific threat level of each enemy in the variable threat condition was chosen randomly, and was held constant across subjects. There were also two recall conditions. In the *free recall* condition, subjects were asked to recall the item locations in any order, whereas in the *serial recall* condition, subjects

paradigm, we have found much stronger primacy in versions that require serial recall than versions that allow free recall.

The only analyses reported here are for the deployment response. We first examined the accuracy of the deployment decision (deployment accuracy), and we next examined what is called here an "end analysis" comparing the distance of subjects' deployment from the first three items to the distance of subjects' deployment from the last three items, to examine the time course of both primacy (anchoring) and recency. This comparison involves the within-subjects factors of end (initial or final) and position (end, second from end, third from end). Both deployment accuracy and the end analysis included the between-subject variables of recall condition (free or serial), threat condition (constant or variable), and counterbalancing subgroup (though no results involving counterbalancing subgroup will be reported here).

Accuracy. For deployment accuracy, the two within-subjects factors were block of 12 trials (one through four), and session (Week 1 or 2) resulting in a 2 x 2 x 6 x 4 x 2 mixed factorial design. For each sequence, there was one optimal deployment location, which was calculated based on all seven enemy locations, for each trial. For the constant threat condition it was the battlefield location with the spatial coordinates closest to the average x and y coordinates of the seven enemy squares. For the variable threat condition, in addition to the spatial locations, the threat level of each enemy was weighted equally in the calculation of the optimal deployment location. Deployment accuracy then was defined as distance of subjects' selected deployment locations from the ideal deployment location, with the distance along the x and y axes recorded separately then averaged for each deployment response. Each subject in a subgroup saw the same 48 seven-item sequences. Feedback showing the optimal deployment location was also given immediately after the subjects' choice, but this feedback displayed an optimal location

based only on spatial location. In this way, we were able to keep the optimal location constant across the experiments reported here, despite increased stimulus complexity.

End analysis. The end analysis examines the influence of each enemy location on the deployment decision. The dependent measure in the deployment analyses was distance, measured in grid squares, from the chosen deployment location to various other points on the screen. For the analysis of deployment accuracy, the dependent measure was distance from the ideal deployment location to the chosen deployment location on each trial. For deployment bias the dependent measure was average distance of chosen deployment location from the spatial position of the item appearing at each serial position, on each trial. For recall the dependent measure was whether or not subjects recalled a given item at the correct serial position, scored as proportion of errors. As mentioned earlier, we counterbalanced the first, second, third, fifth, sixth, and seventh items in each sequence across subjects using a balanced Latin square design. This counterbalancing scheme allowed us to compare three pairs of items that were the same across subjects: The first (Serial Position 1), second (Serial Position 2) and third (Serial Position 3) items were compared to the last (Serial Position 7), second to last (Serial Position 6), and third to last (Serial Position 5) items, respectively. We used this end analysis to assess the withinsubjects variables of end (initial vs final) and position (e.g., Figure 18; end, second from end, third from end). We added these two variables to the variables from the deployment accuracy analysis for a 2 x 2 x 6 x 4 x 2 x 2 x 3 mixed factorial design. A simple effect of end is indicative of primacy relative to recency, or recency relative to primacy. Order effects in decisions based on sequentially arriving information have often been operationalized in this manner in the literature on anchoring in decision-making (Hogarth & Einhorn, 1992). In this analysis, an End x Position interaction describes the more complex pattern of primacy that

consistently obtains in this centroid judgment paradigm, with subjects choosing deployment locations nearer both the beginning and end of each sequence, with this effect of position far more pronounced for the initial three than the final three items presented in each sequence.



Figure 18. End analysis, averaging across all other variables, in Experiment 1. Subjects deploy closer to the first and second items presented in each sequence than any others.

Procedure

After completing a consent form, instructions were presented on the computer screen. Subjects were told: "Read these instructions carefully. As soon as you are sure you understand the experiment, click start to begin. If you have any questions, raise your hand." Subjects then read the instructions, completed the 48 trials, and were debriefed.

The battlefield was represented by a 20 x 20 grid on a grey background. After the brief presentation of a green fixation cross in the center of the grid, seven square targets appeared sequentially, in pseudorandom locations on the screen, with each square representing an enemy location. Each item was presented for .75 s. All seven spatial locations were randomly generated for each trial, unless no single ideal firing location existed for a given set of enemies. When that was the case, new randomly generated enemy locations were substituted one by one

until a single ideal deployment location emerged. Subjects were asked to make their deployment decision, and were then asked to recall the seven locations either in their order of presentation (serial) or in any order (free). Finally, they were asked to recall the threat level of each enemy, in the order that the enemies appeared, by clicking on one of four response options, displaying the four colors of the four threat levels beside the correct location of that enemy. After deployment in the first 24 trials, subjects received feedback in the form of a star showing the optimal deployment location and a square where they fired. They then recalled the seven spatial locations. In the variable threat condition, the actual 7 enemy locations were then re-presented, in order, and subjects were asked to recall the threat level of each enemy. Each new trial was automatically presented upon completion of the previous one.

Results

Accuracy

Again, accuracy in the deployment decision was defined as the distance of subjects' selected deployment locations from the single ideal deployment location. On average, subjects improved from the first to the second block of trials (training phase) and showed similar accuracy on the third block of trials, with accuracy declining during the fourth block of trials, reflected in a main effect of block, F(3,132) = 6.997, p = .0002, MSE = .271. However, this overall pattern is caused by performance during the first session, in which subjects improved from the first to the second block of trials during the first week, but did not improve across blocks the following week (Figure 19), demonstrated by a significant Session x Block interaction, F(3,132) = 8.965, p < .0001, MSE = .226. Subjects in the serial recall condition, but not those in the free recall condition, improved in the accuracy of deployment decisions across blocks in the first week but not in the second, shown in a Session x Block x Recall interaction, F(3,132) =

4.008, p = .0091, MSE = .226. As shown in Figure 19, free recall subjects showed good accuracy (i.e., low distance from ideal) in both sessions, giving them less room for improvement from one week to the next. On average, subjects in the free recall condition deployed their UAVs nearer to the ideal deployment location than those in the serial recall condition (Figure 19, F(1,44) = 5.616, p = .0223, MSE = 4.254). Averaging across all other variables, subjects fired closer to the ideal location during the second session than in the first (Figure 19, F(1,44) = 18.489, p < .0001, MSE = .842), but this effect was driven by the serial recall subjects in the constant threat condition (Figure 20), evidenced by Session x Recall Condition (F(1,44) = 11.872, p = .0013, MSE = .842) and Session x Threat x Recall Condition (F(1,44) = 15.100, p = .0003, MSE = .842) interactions.



Figure 19. Deployment accuracy in Preliminary Experiment 1. Subjects in the serial recall condition improve from the first week to the second more than those in the free recall condition.



Figure 20. Deployment accuracy in Preliminary Experiment 1. Subjects in the serial recall, constant threat condition improve from the first week to the second.

End analysis

As can be seen in Figure 18, subjects deployed their UAVs closer to initial than final list items (F(1,44) = 10.484, p = .0023, MSE = .500), thus demonstrating primacy, and nearer the items toward the ends of the lists, although only toward the front end (e.g., beginning (F(2,88) = 6.544, p = .0022, MSE = .324); that is, this end effect was only evident for the initial items, as evidenced by an End x Position interaction (F(2,88) = 9.459, p = .0002, MSE = .335). The main effect of end was driven by accuracy in the serial recall condition, as demonstrated by a significant End x Recall interaction (Figure 21, F(1,44) = 5.969, p = .0186, MSE = .500). There was also a Session x Position x Recall Condition interaction (Figure 22, F(2,88) = 3.894, p = .0240, MSE = .167), reflecting a more pronounced difference between the end and second to end positions, in serial recall, in Week 1 than all other cases.



Figure 21. End bias in Experiment 1. Serial recall subjects show strong primacy with no recency, whereas free recall subjects show both primacy and recency in their decisions.



Figure 22. End bias in Experiment 1. Serial recall subjects' position bias is reduced from the first to the second week, whereas free recall subjects' position bias remains stable.

Preliminary Experiment 2

We found clear evidence for a primacy bias in subjects' deployment decisions in Preliminary Experiment 1. In Preliminary Experiment 2, we added a level of stimulus complexity. Based on previous suggestions, we expected this extra level of complexity to result in the emergence of recency in the deployment decision (Hogarth & Einhorn, 1992). We also added an instructional debiasing manipulation. We expected to see the same primacy bias in the deployment decision that we found previously, and we expected our instructional debiasing manipulation to attenuate this bias.

Method

Subjects

Twenty-four undergraduates from the University of Colorado Boulder participated for course credit in this experiment in the Spring 2010 semester. They were again assigned by a fixed rotation to one of the six counterbalancing subgroups, with 4 subjects in each subgroup. **Design**

The design of Preliminary Experiment 2 was the same as that of Preliminary Experiment 1, with a few exceptions. First, we did not include a second session 1 week later. We also added another level of complexity in the stimuli: three levels of size, representing the certainty of the intelligence that led to knowledge of that enemy location. As was the case for threat level (color), target sizes (certainty level) were chosen randomly, and were not counterbalanced. Third, we only included free recall of locations for all enemies, followed by serial recall of threat and certainty information. Thus on each trial, subjects (a) saw seven enemies, (b) made their deployment decision with the mouse, (c) recalled the seven enemy locations in any order, and then (d) recalled the seven enemy colors and sizes, in the order that they were presented. Finally, in the single experimental session, we included two experimental halves, training (in which subjects received deployment feedback) and test (in which subjects received no feedback). Before the test phase subjects were shown instructions intended to debias their behavior. The debiasing instructions attempted to draw subjects' attention away from the initially presented item, but fit within the context of the UAV cover story. Toward these ends, subjects were told: "You are engaged in integrating intelligence information from sources collected over time. As is

typical in the dynamic hostile environment, things change, and so earlier arriving information is less reliable than information which has just been received."

Results

Accuracy. For deployment accuracy, again defined as distance of chosen centroid from ideal deployment location, accuracy improved from the first to the second half of the training phase, but decreased from the first to the second half of the test phase, resulting in a main effect of block (Figure 23, F(1,23)=4.100, p=.0546, MSE=.211) and a Phase x Block interaction, (Figure 23, F(1,23)=43.994, p<.0001 MSE=.145).





End Analysis. Figure 24 shows the results of the deployment end analysis for Preliminary Experiment 2. There was a significant effect of position (F(2,46)=14.292, p<.0001, MSE=.658), because on average subjects deployed their UAVs closer to the items presented first and last in the sequences. There was also an End x Position interaction (F(2,46)=4.979. p=.0110, MSE=.389), reflecting overall primacy (or anchoring) in the decision. In addition, there was a significant Phase x End interaction (F(1,23)=4.958, p=.0360, MSE=.574), reflecting the increased influence of the spatial positions of the enemies at serial position 5, 6, and 7 (i.e., in the "final" positions) in the test phase, hence reflecting the intended influence of our instructions.



Figure 24. End analysis in Preliminary Experiment 2. The debiasing instructions increased the influence of the last three items on the decision.

Discussion

Our results support the suggestion that increased stimulus complexity leads to more recency in a decision involving sequentially arriving information (Hogarth & Einhorn, 1992). Recency only emerged at the highest level of stimulus complexity, in Experiment 2, when subjects were asked to weight their centroid judgments by the color and size of each target. In Experiment 2, we also found clear evidence of the efficacy of a simple instructional debiasing manipulation. However, it only reduced the error of recently presented items on the decision. It did not affect primacy.

Appendix B – Contrast of Preliminary and Present Experiments

There are a few limitations of the experiments described in Appendix A, as they were implemented. First, the feedback we provided on the ideal deployment decision was not actually weighted by the level of color or size displayed on the screen, as the initial instructions suggested it would be, nor did it weight the recent items more than the others as the test block instructions suggested. The paradigm was set up this way to control for the influence of feedback between the two conditions, and thus make the comparison between the two instructional conditions clear. However, it makes interpretation of the effect of trial-by-trial feedback in training impossible. In the new experiments we actually weighted the centroid calculation by size and color, described in detail in the main body of the dissertation.

A second major shortcoming of the preliminary experiments was the level of resolution at which we were capable of measuring subject performance within the first version of the experiment. The program was limited to recording whether a subject clicked on a given square of the grid that was visible on the screen. This coarse level of resolution resulted in a bumpy performance metric, based on all-or-none scoring for each item. In the newer version of the experiment used in the main experiments, the location of subject clicks were recorded at a much finer level of resolution -- the pixel coordinates of the locations subjects clicked. This level of detail allows much more sensitive measurement of subject behavior, and thus more complete information regarding subject behavior. Hence we avoided the floor effects that were evident in the recall component in the two preliminary experiments, described in Appendix A.

Finally, the complexity manipulation in the preliminary experiments occurred across the two experiments, which were conducted in two separate semesters. The new Experiment 1 addresses this issue by including all three levels of stimulus complexity that were included in the

preliminary experiments in a single investigation, with subjects assigned to condition by fixed rotation.

Appendix C

Anchoring in Education

Within the span of a semester, a lesson, or even a single concept, students' initial understanding colors what is learned later. Domain-relevant representations that are established during one's introduction to a given concept or domain can serve either to facilitate or impede future learning within that context. In many cases, a given student's initial understanding can be wrong in various ways, but can nonetheless continue to lead to correct answers on tests and assignments until the subject matter students are learning becomes more complex. If the initial representation is *close enough* to optimal, the correction of an inappropriate representation might actually serve to refine and strengthen one's understanding of a given concept or domain (e.g., Clement, Brown, & Zietsman, 1989). But in other cases, suboptimal aspects of an initial representation can actually impede later learning. For example, if students weight earlierpresented information more heavily than information that comes later, but initially presented information is extreme or misleading, building effective representations can first require deconstructing existing representations. Inappropriate anchoring in educational settings can create lasting bad habits and/or misconceptions. These misconceptions can be thought of as bad habits as well, albeit conceptual bad habits.

When these bad habits persist, they are amplified by repeated practice. For example, swinging a golf club in the same manner, repeatedly, reinforces the way it is executed, whether good or bad. In the case of a complex motor skill, self-teaching can result in inappropriate movement patterns, amplified over time, that are far more difficult to change years later, after other representations have been built with misconceptions at the foundation (Smith, Lovatt, & Turner, 2009). Guided exploration, with an appropriate degree of conceptual scaffolding offered

by the coach, is the best way to build appropriate task-relevant representations from the beginning of training. The same is true of complex *cognitive* skills.

As in the case of military and medical decisions, anchoring (the inordinate influence of initially acquired representations within any educational domain) in education can be either good or bad (e.g., Clement et al., 1989; Margetson, 1999; Nickerson, 1998). The initial piece of information a radiologist examines in an x-ray could be the most diagnostic feature in the entire image. In that case an optimal approach would result in a decision biased in the direction of earlier-presented information. Anchoring may even be generally evolutionarily advantageous on average, to the extent that one's early experiences of a thing are indicative of future experiences with that thing. In the classroom, anchoring can also be good, if the initial representation that is established of a thing (in this case an academic domain or topic) is an efficient basis of scaffolding.

In many educational domains, new information regarding a given topic follows from information that has already been established, and so sound initial understandings facilitate later learning, whereas confused initial understandings can be a hindrance (e.g., Clement et al., 1989; Hammer, 1996; Margetson, 1999). The theory of "mastery learning," for example, emphasizes the importance of building appropriate domain-relevant representations early in learning to facilitate later learning (Block, 1971; Block & Burns, 1976; Davis & Sorrell, 1995). Others prefer "preconceptions" to "misconceptions" and discuss less advantageous preconceptions as "brittle," or inflexible for generalization (e.g., Clement et al., 1989). This framing is similar to a distinction from the debiasing literature between "freezing," "moving," and "unfreezing" biases (e.g., Arnott, 2006), with "brittle" representations an analogical descriptor to "frozen" in regard to biases, and the idea of "moving" these representations to an unfrozen state analogous to

97

generalizing or modifying representations. Still other investigators discuss "formation processes" (e.g., Smith et al., 2009). Even on small time scales, the manner in which a task or concept is launched can either facilitate or impede building domain-relevant representations (Ball, Sleep, Boerst, & Bass, 2009).

Externally provided vs. internally created anchors

Concepts are often misunderstood from their first introduction (e.g., Holland et al., 1997; Smith et al., 2009), and such initial misconceptions can be more damaging than those that arise later in learning. At times, misconceptions are explicitly instilled by teachers. These are sometimes a result of teachers discussing a topic in passing, perhaps outside of their areas of expertise. It is difficult for teachers to gauge when exactly students are processing what teachers say - a number of factors influence what environmental inputs any individual is processing at any given time. Perhaps more frequently the limited accuracy of popular pedagogical oversimplifications of complex phenomena might not become a barrier to understanding until much more advanced topics in a given domain.

For example, many chemistry teachers suggest that oil and water do not mix because their molecules repel each other. In fact, the attraction among water molecules is stronger than the attraction between oil and water, although the force of the attraction between oil and water is actually greater than between two molecules of oil. Thus although oil is attracted to water, water is simply more attractive to itself; water molecules are highly polar, whereas oil molecules are not. Although they are difficult to forget, accommodating teacher created misconceptions is possible (Clement et al., 1989). This outcome may be in part because we acquire those misconceptions declaratively, from another individual, and thus the initial acquisition is t2, making continued t2 access to these misconceptions for purposes of modification relatively easy.

Student-developed misconceptions can be more insidious, because like pedagogical conceptual heuristics, they satisfice for a given stage of complexity in material, only revealing themselves as shortcomings much later in learning. However, student-driven misconceptions are more likely to be acquired in a t1 manner than misconceptions instilled by teachers. It is difficult to ever gain conscious access to misconceptions acquired in this manner, and thus they are difficult to uncover and modify. Of course, student-driven misconceptions can also be acquired in a primarily t2 manner, and teacher-driven misconceptions can be acquired in a primarily t1 manner. However, teachers imparting knowledge to students in a traditional lecture situation is usually a declarative interaction, whereas the proportion of t1 and t2 contribution to behavioral control fluctuates over time when students participate in purely self-directed learning, with a greater t1 contribution to control in domains for which students lack appropriate conceptual scaffolding in which to fit new information. A misconceived heuristic for geometry, such as identifying a hypotenuse based on its spatial position on a page, rather than position relative to a right angle, might not actually lead to any noticeable mistakes until a student must solve more complex problems that build on the Pythagorean Theorem. Although self-directed learning takes advantage of both t1 and t2 in learning, learners are less likely to be aware of self-developed misconceptions than those that are explicitly taught. Whether across the timescale of a sentence, a lecture, a semester, or a school career, early misconceptions change the way any subsequent domain-relevant information is organized (e.g., Smith et al., 2009). Such misconceptions can thus reduce the structural integrity of any conceptual scaffolding being built in the mind of the learner.

This distinction follows from the debiasing work of Epley and Gilovich (2005). Under their framing, an important dimension in anchoring is whether an anchor is self-generated or provided externally. Self-generated anchors are more easily influenced by incentives and forewarnings than externally provided anchors. Part of the story has to do with confidence – as in the case of a student who lacks appropriate scaffolding to learn a new concept, subjects with self-generated anchors are less likely to trust them, and are thus more amenable to influence. The same might be true for classroom learning – self-developed misconceptions might be changeable, but only if they can be identified.

Debiasing in Education

A generalizable approach to debiasing could be useful in educational contexts. Automatic biases can lead to suboptimal behavior in the classroom, just as they do in hospitals and on battlefields (e.g., Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007; Holland, Griffiths, & Woodman, 1997). Cognitive biases are thought to arise from t1 cognitive processes, although certain of the heuristics that lead to these biases have also been framed as "cognitive shortcuts" to relieve load on t2. However, any cognitive operation that has been practiced to the point of automaticity can also be thought of based in t1. The generation of educational strategies to influence this t1-based cognition is challenging in part because t1 processes occur much more quickly than the timescale needed for deliberate, verbal processing, and individuals often lack explicit access to implicitly acquired representations. T1 biases can hinder learning, so their identification and amelioration may be an important classroom goal.

The Present Results

The present results might not scale up to classroom situations because these decisions take place on a very short timescale, and they are simple perceptual decisions – a context far

removed from the classroom. However, anchoring has been demonstrated in many situations and across multiple timescales, and similar mechanisms have been proposed across those contexts and timescales (e.g., Chapman, & Johnson, 1999; Hogarth & Einhorn, 1992; Mussweiler & Strack, 1999), so successful strategies to debias anchoring in weighted centroid judgments at short timescales might similarly affect judgments at longer timescales. However, in the present experiments, only manipulations that fundamentally changed the way that subjects approached the task ameliorated decisional bias. Thus avoiding biases from the beginning of learning might be optimal.

Comparison of the results of Experiment 5 to those from the other four experiments suggests that it may be better to avoid establishing initial misconceptions in the first place – the only way to avoid anchoring was to avoid its effect from the outset of learning. There are at least two clear suggestions that arise from this conclusion. First, educators often attempt the transfer of information under an assumption of students as blank slates, with information transfer, storage, and recall as the primary goal. Instead, educators should start by exploring student preconceptions to build appropriate domain-relevant representations from the beginning of learning, with a focus on relevant cognitive skills (e.g., Aron, 2006; Clement et al., 1989, Dewey, 2007; Gurpinar, Musal, Aksakoglu, & Ucku, 2005). Second, learning should be largely student-driven, but appropriate conceptual scaffolding should be offered by the educator to organize experientially learned information frequently.