Information Uptake: Understanding the Impact of New Numerical Information on Quantitative Judgments and Real-World Knowledge

by

Oliver Schweickart

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology
University of Alberta

Abstract

People are constantly exposed to numerical information in their physical and social environments (e.g., food calories, product prices, etc.). Therefore, an important question is when, how, and why this information impacts people's beliefs about the world and their judgments and decisions. Unfortunately, the research literature on this topic provides very different, and at times contradictory, conceptualizations of how quantitative information interacts with real-world knowledge and how it impacts human judgment.

For example, the well-known anchoring and hindsight bias effects observed in quantitative judgment tasks are often portrayed as prime examples of how external information can trigger unconscious and automatic mental processes that inevitably influence people's judgments and beliefs. In contrast, studies on numerical advice-taking highlight people's conservatism when it comes to judgment revision, showing that rejection of new numerical information is quite common, and that people often fail to take advantage of a (generally superior) averaging strategy. Finally, the seeding literature paints a more positive picture, demonstrating that people have a reasonably good ability to draw inductive generalizations from samples of real-world quantitative information.

In this thesis, I propose a unified framework for understanding the above-mentioned phenomena—seeding, advice-taking, anchoring, and hindsight bias—in the context of numerical judgment. This framework asserts (a) that the management of numerical information is generally based on controlled processes,

(b) that people use different response modes (e.g., rejection, adoption, etc.) when they interact with numerical information, (c) that this latter assertion necessitates an analysis on a response mode level to understand the relevant phenomena, and (d) that seeds, advice, and anchors can be conceptualized as numerical information varying along a source credibility continuum.

The central assumptions of this framework were tested in 4 experiments. In Experiments 1-3, a hybrid advice-taking/seeding paradigm was used where participants first generated population estimates for a set of countries, then were exposed to numerical information for a subset of these countries, and finally provided a second set of estimates for all countries. These experiments revealed (a) that information utilization varied as a function of the source credibility of the information, and that the aggregate source credibility effect was driven by the adoption rate; (b) that irrespective of the source credibility level, the provided numerical information elicited transfer; however the presence of transfer was contingent on prior information utilization; (c) that informational context—defined as the numerical information made accessible to the decision maker during the target judgment—impacted the rejection rate, specifically the accessibility of one's prior estimate increased the likelihood of rejection.

The primary objective of Experiment 4 was to test competing predictions between models of hindsight bias that link the effect to automatic processes underlying knowledge revision, and the current framework of controlled information processing. This experiment employed a hybrid advice-taking/hindsight bias paradigm where participants first answered a heterogeneous set of estimation questions, and then were given the opportunity to revise their judgment in response to numerical advice. Finally, participants had to recall their initial estimates. Counter to the predictions of automatic accounts, the results of this study indicated that the presence of hindsight bias depended on the prior utilization of advice. In other words, if advice was rejected, no hindsight bias emerged in the subsequent recall task.

In sum, the results of these experiments are consistent with the proposed framework that highlights the role of controlled processes in quantitative judgments under uncertainty. I end by discussing implications of these findings for understanding numerical judgment in real-world knowledge domains.

Preface

This thesis is an original work by Oliver Schweickart. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Names "Information Uptake", No. Pro00016829, 08/23/2010; “Information Uptake Project”, No. Pro00025804, 10/13/2011; "Making and Judging Numerical Estimates", No. Pro00034351, 10/11/2012.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Norman Brown, for the continuous support during my graduate studies and research. His encouragement, broad expertise, availability to discuss "quick questions" at all times (even though these questions often required lengthy answers), and his unique and brilliant sense of humor made life as a graduate student so much more enriching and enjoyable. Thanks for having been such an excellent *Doktorvater*.

I would also like to thank my present and former committee members, professors Alinda Friedman, Peter Dixon, and Tom Spalding, as well as my external examiners, professors Gerald Häubl and Hartmut Blank, for their superb comments, insights, and suggestions. Furthermore, I want to extend my appreciation to Professor Weimin Mou for chairing my defense.

Parts of this research were supported by a SSHRC Insight Grant awarded to my advisor Dr. Brown, and I'm grateful for the funding I received through this grant.

During my stay at the University of Alberta, the members of the Brown Lab quickly turned from colleagues into valued friends. Thanks to Tuğba Uzer, Connie Svob, Cory Tam, and Amelia Shi for creating such a supportive and friendly environment. I'm also grateful to the research assistants who have helped me with data collection in various research projects throughout the years.

Last but not least, a special thank you goes to my family for their unwavering support and continuous encouragement.

Table of Contents

List of Tables

List of Figures

1 Introduction

A Starbucks Venti Double Chocolaty Chip Frappuccino has 520 calories, the same as a McDonald's Quarter Pounder with Cheese.

Worldwide, 780 million people lack access to clean water (that is more than 2.5 times the United States population).

There are reportedly 114 million handguns in civilian possession in the United States.[1]

1.1 How Does Quantitative Information Impact People's Judgments and Beliefs?

Numerical information is ubiquitous: from numeric labels and nutritional information, to health statistics, public opinion polls, and product prices; numbers are everywhere. Hence, people constantly have to cope with numerical information in their physical and social environments. An important question is therefore when, how, and why this information impacts people's beliefs about the world and their judgments and decisions.

[1] Numerical facts obtained from: http://www.starbucks.com/menu/drinks/ frappuccino- blended-beverages/double-chocolaty-chip-frappuccino-blended-creme#size=11002679&milk=67&whip=125; http://www.mcdonalds.com/us/en/food/product_nutrition.sandwiches.286.quarter-pounder-with-cheese.html; http://water.org/water-crisis/water-facts/water/; http://www.gunpolicy.org/firearms/region/united-states

The present research is directed at understanding the psychological underpinnings of *quantitative literacy*—the "ability to make reasoned decisions using general world knowledge and fundamental mathematics in authentic, everyday circumstances" (Wiest, Higgins, & Frost, 2007, p. 48)—and to understand these in the context of socially and economically relevant content domains (e.g., national populations, GDPs, product prices, etc.). Particularly in today's Information Age, being quantitatively literate is important, if not essential, to understand complex political and social issues, to assess potential risks and benefits, and to make informed financial, health, or voting decisions (Reyna, Nelson, Han, & Diekmann, 2009). For example, quantitative literacy is critical for policy makers to have accurate and unbiased beliefs about important economic, environmental, or health issues. It is critical for the electorate to participate in public discourse and to avoid being potentially misled by politicians or marketing campaigns, and it is critical for the individual to be an informed voter, consumer, and decision maker.

The question of how quantitative information interacts with real-world knowledge and how it impacts human judgment has been studied in several sub-fields of psychology (reviewed below). However, this research literature is poorly integrated and has provided very different, and at times contradictory, conceptualizations of (a) how easily people are influenced by new information, (b) what types of judgmental biases the information evokes (and what psychological mechanisms underlie these biases), and (c) what the information's impact and long-term consequences are for the underlying knowledge base. Four research literatures that directly speak to these issues are *seeding the knowledge base* (e.g., Brown, 2002; Brown & Siegler, 1993), *advice taking* (e.g., Yaniv, 2004a; Bonaccio & Dalal, 2006), *judgmental anchoring* (e.g., Tversky & Kahneman, 1974; Furnham, & Boo, 2011), and *hindsight bias* (e.g., Fischhoff, 1975; Fischhoff & Beyth, 1975; Roese & Vohs, 2012). All these literatures share a similar experimental paradigm (see Table 1.1) where participants are exposed to numerical information and then, in a subsequent task, the impact of this information on judgment and/or the underlying knowledge base is measured. The overarching goal of the present research project is to relate findings from these different literatures and understand their psychological underpinnings within a common framework.

Table 1.1 Experimental paradigms related to information uptake

Experimental paradigm	Pre-information phase	Information exposure	Post-information phase
Seeding the knowledge base	Generate Est 1 *Seed* *Transfer*	Learn *Seed*	Generate Est 2 *Seed* *Transfer*
Advice taking (JAS system)	Generate Est 1	Learn *Advice*	Generate Est 2
Standard anchoring	-----	Assess *Anchor*	Generate Est
Hindsight bias (memory paradigm)	Generate Est	Learn *Outcome*	Recall Est

Note. Est = estimate.

In the remainder of this chapter, I will first provide a brief review of the major phenomena reported in these four literatures and highlight the different conceptualizations that were developed within each of these paradigms as to how new information impacts judgment and real-world knowledge. Then, I will outline a theoretical framework designed to integrate these phenomena and to understand their underlying psychological mechanisms. Finally, I will provide an overview of the aspects of this framework that were tested in the research reported in this thesis.

1.2 Information vs. Contamination

The major schism in the above-mentioned literatures pertains to the level of control attributed to the decision maker when interacting with the new information. In the seeding and advice taking literatures the decision maker is typically conceptualized as someone who assesses the new information and uses it, if it is deemed relevant (even though the ways in which the information is integrated and used might not be optimal by normative standards). That is, here the dominant view is that of controlled information processing. In contrast, research on anchoring and hindsight bias suggests that the decision maker's judgments and knowledge base are easily, and unavoidably, altered (or "contaminated") by new information. In these literatures an important idea is that the exposure to new information in the respective tasks triggers mental processes that are automatic and unconscious (e.g., Wilson & Brekke, 1994; references below; for a critical review see Newell & Shanks, 2014a, 2014b), and that specific debiasing techniques are required to counteract this influence (Arkes, 1991; Fischhoff, 1977; Mussweiler, Strack, & Pfeiffer, 2000).

Thus, in their relative isolation, researchers studying these different, yet related, phenomena have arrived at very different conclusions regarding the relative contributions of automatic and controlled processes in (quantitative) judgment (Evans, 2008; Kahneman, 2011; Newell & Shanks, 2014a, 2014b).

Seeding the knowledge base. In the standard seeding paradigm, people are first asked to generate a set of estimates for items sampled from a real-world knowledge domain (e.g., populations of countries). Then, they learn the actual values for a subset of these items, the so-called *seed* fact(s), and are subsequently asked to generate a second set of estimates for all items. This paradigm has been successfully used to study the processes involved in generating real-world estimates (Brown, 2002; Brown, Cui, & Gordon, 2002; Brown & Siegler, 1993, 1996, 2001; LaVoie, Bourne, & Healy, 2002; Smith & Windschitl, 2015; Wohldmann, 2015), as well as to illuminate the organization of complex, real-world knowledge (Friedman & Brown, 2000a, 2000b). One of the key findings in this literature is that people draw inductive generalizations from the provided numerical information. That is, after learning the seed facts, accuracy of the non-seeded items (i.e., transfer items) also improves, which suggests that people use the provided information to update their metric beliefs about the target domain in general. Seeding effects have been shown to be long-lasting—at least 4 months—

(Brown & Siegler, 1996), thus the findings obtained from seeding the knowledge base point to the potential benefits of learning new (and accurate) numerical information.

Advice taking. A structurally similar paradigm is employed in the advice taking field to study how people revise judgments in response to advice. In this paradigm (also called the judge-advisor system [JAS], Sniezek & Buckley, 1995), participants first answer a set of quantitative, real-world knowledge questions taken from different domains (e.g., historical dates, calories, etc.). Then, the participant is presented with each question again, and is shown his or her prior estimate as well as a piece of numerical advice (typically one target estimate provided by somebody else), and has to generate a final estimate. This research has focused on understanding how different variables related to the advisor, the advice, or the judge (i.e., the participant) influences the extent to which participants' final judgments shift towards the advice (e.g., Gino & Moore, 2007; Gino & Schweitzer, 2008; Gino, Shang, & Croson, 2009; Harries, Yaniv, & Harvey, 2004; Harvey & Fischer, 1997; Harvey, Harries, & Fischer, 2000; Koehler & Beauregard, 2006; Lim & O'Connor, 1995; Soll & Mannes, 2011; Tost, Gino, & Larrick, 2012; Yaniv, 1997; Yaniv, 2004a, 2004b; Yaniv & Choshen-Hillel, 2012; Yaniv & Choshen-Hillel, Milyavsky, 2009; Yaniv & Kleinberger, 2000; Yaniv & Milyavsky, 2007; for a review see Bonaccio & Dalal, 2006).

Here, the central finding is that people generally weigh their own opinion more heavily than the advisor's opinion, such that the final aggregate estimate is typically shifted towards the advice by about 30% (Yaniv & Kleinberger, 2000). This *egocentric advice discounting* is problematic because in most environments an equal weighting of initial estimate and advisor's estimate (i.e., averaging) would be a superior aggregation policy in order to maximize accuracy gains (Larrick & Soll, 2006; Soll & Larrick, 2009). Thus, egocentric advice discounting prevents people from fully benefitting from the averaging principle that underlies the wisdom of crowds—namely that the averaging of individual estimates provided by a group of independent judges results in a more accurate target estimate than the estimate of a person with an average level of accuracy (Herzog & Hertwig, 2009; Larrick, Mannes, & Soll, 2012; Surowiecki, 2005; Vul, & Pashler, 2008).

Furthermore, a decomposition of the aggregate advice-taking effect suggests that the overall shift of about 30% is an averaging artifact (Soll & Larrick, 2009). Namely, few people actually provide final estimates that are shifted 30% towards the advice; instead people seem to respond to the advice in different ways, some stick to their initial estimates, some adopt the advisor's estimate, and others provide estimates that fall somewhere between the two (with a modest preference for averaging). Given the fact that for a subset responses—overall, typically between 30-40%—people rejected the advisor's estimate entirely, strongly suggests that judgment revision is a controlled process.

Anchoring. Probably the best-known phenomenon (and paradigm) in the area of quantitative judgment under uncertainty is anchoring. The standard anchoring paradigm (Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1974) consists of a two-stage procedure during which participants first determine whether the value of an unknown quantity is greater or less than a (supposedly) arbitrary number (e.g., “Is the height of Mount Everest greater than or less than 45,500 feet?”); then, they provide an estimate of that target quantity (e.g., “What is the height of Mount Everest [in feet]?”). Typically, two anchor values are used, one high (e.g., the 85th percentile of a distribution of unanchored estimates) and one low (e.g., the 15th percentile of a distribution of unanchored estimates), and the central finding is that the aggregate target estimates are biased in the direction of the respective anchors.

Anchoring is frequently presented as a prime example of the susceptibility of human judgment to irrelevant information. This is because anchor values are generally either discredited at the outset (e.g., by telling subjects that these values were randomly generated), or, in other cases, are generated by the participant from judgment-irrelevant information such as the participant’s phone number (Ariely, Loewenstein, & Prelec, 2003; Chapman & Johnson, 1999; Englich, Mussweiler, & Strack, 2006; Russo & Schoemaker, 1989); yet, these values still exert an influence on judgment.

Anchoring effects are robust and have even been demonstrated to occur, for example, when participants are warned in advance about the existence of anchoring (Epley & Gilovich, 2005; Wilson, Houston, Etling, & Brekke, 1996), when the anchor value is presented subliminally (Mussweiler & Englich, 2005; Reitsma-van Rooijen, & Daamen, 2006), or when task-irrelevant incidental numbers in the subject's environment are manipulated (Critcher & Gilovich, 2008; Wilson et al., 1996; but see Brewer & Chapman, 2002; Matthews, 2011). This robustness, and the difficulty to debias anchoring, have led several researchers to conclude that the effect arises in part, or entirely, due to automatic, priming-based[2] processes (e.g., Blankenship, Wegener, Petty, Detweiler-Bedell, & Macy, 2008; Chapman & Johnson, 1994, 1999; Chaxel, 2014; Critcher & Gilovich, 2008; Mussweiler & Strack, 1999; Oppenheimer, LeBoeuf, & Brewer, 2008; Strack & Mussweiler, 1997; Wilson et al., 1996; Wong & Kwong, 2000; see Newell & Shanks, 2014a; Schweickart, Tam, & Brown, 2014 for critical reviews). From this perspective, anchoring effects are seen as "enigmatic" and "inevitable" (Strack & Mussweiler, 1997). And even Daniel Kahneman—the co-inventor of the paradigm—recently concluded that "any number that you are asked to consider as a possible solution to an estimation problem will induce an anchoring effect" (Kahneman, 2011, p. 120).

---

[2]Different types of priming effects have been suggested to underlie anchoring: numeric priming (Wong & Kwong, 2000; Wilson et al., 1996), semantic priming (Mussweiler & Strack, 1999; Chapman & Johnson, 1999), and magnitude priming (Oppenheimer et al., 2008).

Interestingly, the main finding in both the advice-taking and anchoring literatures consists of an aggregate estimate being shifted towards a numerical reference value (advice and anchor, respectively). However, in the former case researchers identify the bias—egocentric advice discounting—as one where people do not move far enough towards the numerical reference value (to take advantage of averaging), and in the latter case, any movement towards the reference value is viewed as a bias—namely anchoring. Furthermore, the finding obtained in advice-taking studies that, overall, advice is rejected on about a third of trials appears inconsistent with the view that numerical information, due to triggering automatic, priming-based processes, will inevitably exert an influence on quantitative judgments under uncertainty. In other words, the current state of the literature creates the paradoxical situation where one group of researchers claims that people frequently reject potentially relevant information (as shown in advice-taking studies), but another group of researchers claims that people are *not* able to fend off the influence of *ir*relevant information (as suggested by proponents of priming-based accounts of anchoring).

Hindsight bias. Research on hindsight bias often incorporates quantitative estimation questions to understand how outcome feedback (i.e., learning the answer to a question) affects the ability to recall one's prior estimate, an estimate that was produced before the outcome was known (Christensen-Szalanski & Willham, 1991; Hawkins & Hastie, 1990; Pohl, 2004). In the memory hindsight bias paradigm, participants first provide a set of estimates for relatively difficult estimation problems. Typically after a delay of one week, participants are provided with the solutions for half of the questions (the other half serves as control). Then, they have to recall their prior estimates that they provided one week earlier. The standard finding is that the presentation of outcome feedback causes people's recollections of their initial estimates to be biased in the direction of the outcome.

Because the present focus lies on understanding the relationship between the processing of numerical information and knowledge updating, the review of the hindsight bias literature is very selective and will only cover research specifically related to the task at hand (see Blank & Nestler, 2007; Blank, Nestler, von Collani, & Fischer, 2008; Nestler, Blank, & Egloff, 2010; Roese & Vohs, 2012 for more recent overviews and theoretical contributions).

One interpretation of hindsight bias is that it is a by-product of knowledge updating (Hoffrage, Hertwig & Gigerenzer, 2000; Pohl, Eisenhauer & Hardt, 2003). According to this view, learning the solution to a question triggers automatic processes that alter the structure and/or content of the underlying knowledge base. When people subsequently try to recall their initial estimate and fail to recollect it, they will make an attempt to reconstruct the estimate (i.e., regenerate[3] an estimate); however this reconstruction process will be based on an altered knowledge base. This, in effect, will cause the reconstructed estimate to be

[3]An alternative view is that the reconstruction process involves an adjustment-based process (e.g., Ash, 2009)

biased in the direction of the outcome. From this perspective, hindsight bias can be viewed as adaptive, because it keeps the knowledge base up-to-date.

One of the central questions regarding hindsight bias is whether it is the result of automatic or controlled processes (Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988). Because the effect is robust and debiasing attempts, such as forewarnings, have failed to reduce the effect, several researchers have concluded that hindsight bias has an automatic component (e.g., Pohl & Hell, 1996; but see Pohl, 1998). In fact, the most advanced theoretical account of hindsight bias in quantitative estimation (Pohl et al. 2003) asserts that anchoring and hindsight bias share the same underlying mechanism, namely the selective activation of anchor/outcome-consistent knowledge. Because selective accessibility is an automatic process (Mussweiler & Strack, 1999), it suggests that both anchoring and hindsight bias should be difficult to eliminate.

1.3 A Simple Framework for Understanding Information Uptake

At present, the seeding, advice-taking, anchoring, and hindsight bias literatures are not very well integrated, and only few attempts have been made to relate them directly (for notable exceptions see Brown & Siegler, 2001; Hardt & Pohl, 2003; Pohl, 1998; Pohl, Eisenhauer, & Hardt, 2003). Here, I want to argue that our understanding of these phenomena can be furthered by viewing them as "case studies" of (numerical) information uptake. That is, all these research paradigms address a common basic issue—How, why, and when does numerical information impact judgment and the content and/or structure of the underlying knowledge base?—and the different phenomena should therefore be conceptualized within a unified framework. In what follows, I want to outline such a framework.

Response modes and cognitive control. The advice-taking literature (as reviewed above) indicates that people react differently to the information that is provided to them (Soll & Larrick, 2009). That is, some people stick to their prior beliefs, some people adopt the new information, and some people combine the new information with their prior estimate in some way (with averaging being a special case), resulting in a final judgment that is different from both the initial estimate and the information. Given that (a) not everyone is influenced by the provided information (i.e., the rejection rate is non-zero), (b) that people seem to use different *response modes* (e.g., rejection, adoption, combining) when they are exposed to new information, and (c) that the relative frequency with which these different response modes are used varies in predictable ways (e.g., the frequency of adoption and rejection depends on the relative domain expertise of the advisor and judge; Soll & Larrick, 2009), suggest that people have a great deal of control over managing numerical information in their environment. From this perspective, the integration of new information and the revision of the knowledge base is seen as an active process requiring a certain amount of cognitive control, as opposed to one that consists of passive encoding driven by automatic processing[4]. From this point of view, information that is assessed as irrelevant for the quantitative target judgment should therefore not elicit anchoring effects, and should not be integrated into the knowledge base (i.e., this information should neither evoke hindsight bias nor seeding effects).

Of course, these predictions are only testable if one moves the analysis of judgments from an aggregate level to a response-based level. If people use different response modes when encountering new information, then the aggregate effect (e.g., percent shift towards the information) reflects an averaging artifact due to the use of different response, and thus masks the underlying psychological processes (Siegler, 1987).

Source credibility continuum. One of the main differences between the four experimental paradigms is the credibility of the numerical information. In the seeding and hindsight bias tasks, the participant is given the actual values for the target items. Therefore, this information should be viewed as highly credible. In the advice taking task, the credibility of the advice is typically left unspecified, thus adding another dimension of uncertainty to the judgment. Finally, in the anchoring task the information is discredited at the outset, thus anchors are situated at the lowest level of credibility. (Of course, simply because a random

[4]Thus, this framework is an attempt to explain as many of the phenomena as possible with a controlled account before resorting to dual-process theories which are less parsimonious and often prove difficult to test and falsify.

number has a low credibility to be an accurate answer to an estimation problem, does not mean no on views it as plausible.)

Thus, the different paradigms can be conceptualized as varying along a source credibility continuum (Pohl, 1998). Source credibility has been shown to be a potent predictor of changes in attitudes, beliefs, or evaluations (Birnbaum & Stegner, 1979; Birnbaum, Wong, & Wong, 1976; Petty & Cacioppo, 1986; Toglia, Ross, Ceci, & Hembrooke, 1992; but see Plous, 1989; Switzer & Sniezek, 1991), and given the current framework, it should also predict the use of different response modes in information uptake (Soll & Larrick, 2009).

Estimation strategies. Assuming that people use different response modes in the task at hand, the final judgment will likely be based on different types of cognitive processes. Understanding which estimation strategy or strategies map onto which response modes is therefore another important issue. For example, adopting an advisor's estimate will involve less computation than generating an estimate that combines the initial estimate and the advisor's estimate. The latter could be generated, for example, by computing a (weighted) average, or by using an adjustment-based strategy, where one "anchors" onto, say, the initial estimate and adjusts towards the advisor's estimate until one reaches a subjectively reasonable estimate (Epley & Gilovich, 2001, 2004, 2006; Lim & O'Connor, 1995; Tversky & Kahneman, 1974).

Besides these numerically-based strategies, a second important estimation mode is *ordinal conversion* (Brown, 2002; Brown & Siegler, 1993). This strategy makes use of two types of knowledge, *metric* knowledge (knowledge/beliefs about statistical properties [mean, range, distribution] of the target dimension) and *mapping* knowledge (knowledge/beliefs about the relative ordering of items along the target dimension). Thus, when generating an estimate, the relative magnitude of a target item is first determined (e.g., Switzerland has a small population), and then a value from the appropriate portion of the response range is selected (e.g., 8 million). There is strong evidence that this strategy is commonly used in the seeding task, but might also play a role in the anchoring and hindsight bias. In anchoring, ordinal conversion might be used when a question involves an extremely implausible anchor (e.g., Is the number of career goals scored by Wayne Gretzky greater than or less than 2?), and people generate an estimate from scratch rather than using the anchor (Chapman & Johnson, 1994; Wegener, Petty, Detweiler-Bedell, & Jarvis, 2001). In the hindsight bias task, ordinal conversion might be used to reconstruct an initial estimate when it cannot be recollected (see Pohl et al., 2003 for a different conceptualization of the reconstruction process). Thus, understanding the mapping of estimation strategy/strategies onto response modes is another important component in understanding the phenomena associated with information uptake.

1.4 Overview of the Present Research Project

The research reported in this thesis integrates the experimental paradigms of seeding, advice-taking, and hindsight bias and tests specific predictions derived from the framework outlined above. Experimental test of anchoring predictions will not be included in the empirical part, but key findings will be briefly summarized in the final chapter.

In Chapter 2, I report an experimental study that links the seeding and advice-taking paradigms. That is, participants were “seeded” with numerical advice. This study investigated how different levels of source credibility impact information uptake, and addressed the following questions: (1) How does the credibility of the source of the information (e.g., almanac value vs. estimate from another student with accuracy level X) affect the use of different response modes? (2) Does advice transfer? (3) If so, does the amount of transfer vary as a function of source credibility? (4) Given that people frequently discount advice, will this also prevent knowledge revision?

In Chapter 3, I examine if and how the choice of different response modes interacts with task characteristics. Specifically, I report an experiment that tests how the informational context at the time of the target judgment (i.e., what numerical information is displayed to the decision maker: the initial estimate and the advice vs. the advice only) impacts the relative use of different response modes. This manipulation also allows to test predictions of different accounts of egocentric advice discounting. The experiment employs the same hybrid seeding/advice-taking paradigm that was used in Experiment 1. Furthermore, it adds another level of source credibility and also examines the transfer of advice.

In Chapter 4, I further examine how the informational context impacts advice taking. An experiment is reported that in part replicates and in part extends Experiment 2 by including two additional contexts; one where only the initial estimate is displayed, and one where no information is displayed during the target judgment. At the end of this chapter I provide a summary of the response mode analyses from Experiment 1-3 to illustrate the usefulness of this type of analysis for understanding the effects of different types of variables that impact information uptake.

In Chapter 5, I report a study that integrates the advice-taking and the hindsight bias memory paradigms. That is, participants first completed the standard advice taking task and then tried to recall their pre-advice estimates. Here, the primary goal was to determine whether a subject's response to the numerical advice would predict the magnitude of the hindsight bias for the recalled initial estimates. The most important question being whether rejected advice would still be integrated into the knowledge base and would thus interfere with the recall of prior beliefs.

Finally, in Chapter 6 I provide a brief summary of the overall findings reported in the thesis and discuss their implications for the understanding of numerical judgment, both in terms of bias as well as potential ways of improvement. Furthermore, I will discuss some of the limitations of the present studies and will point out avenues for future research.

2 Experiment 1 – Seeding the Knowledge Base With Numerical Advice

2.1 Introduction

People are constantly exposed to numerical information in their physical and social environments. This information might be directly relevant for a target judgment (e.g., if your friend happens to know how many calories are in a McChicken sandwich, it might help you to estimate the number of calories in a Big Mac), or it might be completely unrelated to the target judgment (e.g., if, while ordering, you happen to overhear a conversation between restaurant employees that today is the owner's mother's 100th birthday). But even when exposed to potentially task-relevant numerical information, decision makers still find themselves in a state of uncertainty as to how accurate this information is (e.g., how certain are you that your friend's knowledge of caloric values of fast food menu items is accurate?).

The goal of the present study is to understand how task-relevant numerical information associated with different levels of source credibility impacts numerical judgments. A second goal is to understand if and how this information gets integrated into the decision maker's knowledge base. Two research literatures, at the present completely isolated from one another, directly speak to these issues; the *seeding* literature (e.g., Brown, 2002; Brown et al., 2002; Friedman & Brown, 2000a, 2000b; Brown & Siegler, 1993, 1996, 2001; LaVoie et al., 2002; Smith & Windschitl, 2015; Wohldmann, 2015), and the *advice-taking* literature (e.g., Gino & Moore, 2007; Gino & Schweitzer, 2008; Gino et al., 2009; Harries et al., 2004; Harvey & Fischer, 1997; Harvey et al., 2000; Koehler & Beauregard, 2006; Larrick & Soll, 2006; Lim & O'Connor, 1995; Sniezek & Buckley, 1995; Soll & Larrick, 2009; Soll & Mannes, 2011; Tost et al., 2012; Yaniv, 1997; Yaniv, 2004a, 2004b; Yaniv & Choshen-Hillel, 2012; Yaniv et al., 2009; Yaniv & Kleinberger, 2000; Yaniv & Milyavsky, 2007; for a review see Bonaccio & Dalal, 2006).

In both, researcher rely on a 3-phase paradigm where participants first generate numerical estimates of unknown target quantities, then receive numerical information, and finally provide a second set of estimates. In the standard *seeding* task, participants first provide estimates for items drawn from a single knowledge domain (e.g., populations of different countries). Then, the participant learns the true values for a subset of items, and provides final estimates on both the feedback items (seeds) as well as the non-feedback items (transfer items).

Researchers have employed this paradigm to demonstrate, among other things, the dissociation between two types of knowledge relevant in real-world estimation. First, *metric* knowledge refers to knowledge or beliefs about statistical properties of the target dimension (e.g., the mean, range, shape of distribution, etc.). The second type of knowledge is *mapping* knowledge, knowledge or beliefs about the relative ordering of items along the target dimension (e.g., Canada has a smaller population than the United States, which has a smaller population than China; Brown & Siegler, 1993). This dissociation becomes apparent in so-called *seeding* effects. Seeding effects refer to the finding that after learning the true values for the seed items, the accuracy of estimates for the transfer items, on average, also improves; however, this is generally only reflected in measures of metric knowledge, but not in measures of mapping knowledge. This suggests that people use the sample of seed facts to update their beliefs about the metric properties of the target domain. This set of updated metric beliefs is then used to generate the final estimates, which, on average, also decreases the metric error of the transfer item estimates. Thus, seeding effects signify people's ability to draw inductive generalizations from numerical information and also reveal parts of the organization of knowledge involved in real-world estimation.

The advice-taking paradigm differs from the seeding paradigm in four ways. First, the accuracy of the numerical information (i.e., advice) is unknown to the participant and has to be inferred. Second, no transfer items are included. Third, generally one (but occasionally also multiple) estimate(s) are offered as advice for each question. Fourth, both the participant's initial estimate and the advisor's estimate are accessible to the participant when he or she is given the chance to revise the initial judgment.

The research based on this paradigm indicates that people have a tendency to place a greater weight on their initial estimate than the advisor's estimate, such that the final response is biased towards their own beliefs. This finding is known as *egocentric advice discounting*, and reflects an aggregate shift of participants' initial estimates towards the advice of about 30% (Yaniv & Kleinberger, 2000).

However, the size of this aggregate shift has been shown to be modulated by various factors pertaining to the advisor, the task, the advice, and the judge. For example, the aggregate shift has been shown to be influenced (a) by characteristics of the advisor such as expertise (Harvey & Fischer, 1997; Soll & Larrick, 2009; Yaniv & Kleinberger, 2000), reputation (Yaniv & Kleinberger, 2000), similarity to the judge (Gino et al., 2009), confidence (Sniezek & Buckley, 1995; Soll & Larrick, 2009); (b) by characteristics of the task, such as when the task is difficult (Gino et al., 2009), important (Harvey & Fischer, 1997), or when the advice has to be paid for (Gino, 2008); (c) by characteristics of the numerical advice, for example, how distant it is from the judge's initial estimate (Yaniv, 2004a; Ravazzolo & Røisland, 2011; Schultze, Rakotoarisoa, & Schulz-Hardt, 2015); and (d) by characteristics of the judge, such as the subjective sense of power (Tost et al., 2009), emotional state (Gino & Schweitzer, 2008), or expertise (Soll & Larrick, 2009). The fact that these variables modulate the aggregate discounting effect suggests that both the assessment and the use of advice is in large part driven by controlled processes. In brief, advice utilization might not be optimal (see below), but it is intentional.

What underlies egocentric advice discounting? Three main explanations have been discussed in the advice-taking literature (Bonaccio & Dalal, 2006; Soll & Mannes, 2011). The first account attributes the effect to a differential access to the reasons and justifications underlying the judge's and the advisor's estimates (Yaniv, 2004a; Yaniv & Kleinberger, 2000). That is, while people know why they came up with the numerical estimate that they did, they do not have access to this information for the advisor's estimate. In effect, this imbalance in knowledge accessibility biases people towards their initial beliefs. The second explanation asserts that people have biased beliefs about their own ability and/or knowledge (e.g., Dunning, Heath, & Suls, 2004; Kruger, 1999). From this perspective, egocentric advice discounting reflects an above-average-effect, where people believe that their own ability or knowledge is better (more accurate) than those of the average person. This belief, in effect, leads to the overweighting of one's own estimate relative to the advisor's. Finally, the third explanation attributes egocentric advice discounting to the judgment process. According to this view, underweighting of advice arises because people use an anchor-and-adjustment strategy when they revise their final judgment (Lim & O'Connor, 1995). People use their initial estimate as the starting point and incrementally adjust towards the advice; however this adjustment process is insufficient (Epley & Gilovich, 2005; Tversky & Kahneman, 1974). That is, the adjustment process terminates typically after about 30% of the distance between initial estimate and advisor's estimate, thus creating the aggregate effect of egocentric advice discounting.

There is strong evidence that suggests that the last account is probably incorrect. When the aggregate effect of egocentric advice discounting is decomposed—that is, the distribution of the weight placed on the initial estimate is plotted—it becomes evident that only very few of the final estimates shifted towards the advice by about 30% (Soll & Larrick, 2009). Instead, the distribution of the weight placed on the initial estimate (weight on self) is multi-modal. That is, for some people the weight is 1 (reflecting that they did not change their beliefs in response to the advice), for some the weight is 0 (reflecting that they adopted the advisor's estimate), and for some the weight is somewhere between 0 and 1, with an additional mode at 0.5 (averaging). This finding is important for several reasons. First, it suggests that people can and do react differently to numerical information in their social environment and that the aggregate shift of roughly 30% is an averaging artifact (Siegler, 1987). Second, the fact that a weight of 1 (i.e., rejection of advice) was observed in roughly 35% of trials suggests that people have a great deal of control over how they manage numerical information. As well, this finding is inconsistent with the dominant theoretical account of a related numerical judgment phenomenon—anchoring (e.g. Strack & Mussweiler, 1997)—which asserts that the exposure to and evaluation of numerical information will trigger automatic processes that inevitably influence the target judgment.

From a normative point of view, egocentric advice discounting is problematic because it results in a suboptimal aggregation of the information in the task at hand. In many environments, an equal weighing of the judge's own estimate and the advisor's estimate would result in a greater overall accuracy (Larrick & Soll, 2006; Soll & Larrick, 2009; Soll & Mannes, 2011). However, when making their final judgment, people tend to prefer to either use their own estimate (i.e., reject the advice) or use the advisor's estimate (adopt the advice).

This is true, even if the judgment environment strongly favors an averaging strategy. However, when people are asked to combine the estimates of two other people to arrive at the best possible judgment, people are much more likely to average the two (Soll & Mannes, 2011). On the basis of these findings, Soll and colleagues (Soll & Larrick, 2009; Soll & Mannes, 2011) concluded that people use two different strategies when they revise their judgment, choosing (i.e, either adopting the advice or rejecting it), and averaging (i.e., equal weighting of advice and initial estimate).

However, this conclusion might be problematic. Even though the distribution of the weight on self was found to be tri-modal (with modes at 0 [adoption], 0.5 [averaging], and 1 [rejection]), the criterion for classifying a response as averaging was fairly liberal (weight on self between 0.4-0.6), yet still roughly 30% of responses could not be classified as either choosing or averaging.

This suggests that the choosing and averaging strategies do not capture the full spectrum of people's responses. Therefore, I adopt a more descriptive approach and categorize people's response into 5 different categories (see Table 2.1). *Adoption* refers to responses where the final judgment (roughly) equals the presented numerical information, and *rejection* refers to responses where the final judgment roughly equals a person's initial estimate. *Assimilation* denotes those responses that result in an estimate in between the initial estimate and the presented information. *Overshoot* occurs when a person's final estimate moves beyond the information given, and *contrast* occurs when the final estimate moves away from both the person's initial estimate and the provided numerical information. Note that these five *response modes* are only used as descriptive terms to classify the different responses in the task at hand. I remain agnostic about the processes and motivations underlying each response mode.

The present study was designed to answer the following research questions. First, how does source credibility impact the use of different response modes? Prior research within the advice-taking paradigm has shown that, on an aggregate level, the weight on self decreases as the credibility of the source increases (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). But this aggregate effect could emerge in different ways. First, people might combine the two estimates and, with increasing source credibility, successively increase the weight placed on the advisor's estimate. Another possibility is that source credibility effect is primarily driven by the adoption rate. That is, as source credibility increases, the frequency of adopting the advice will increase. Soll and Larrick (2009) obtained results consistent with the latter view, but the issue has not been explored systematically.

A second question was if people draw inductive generalizations from numerical advice in the same way as they do from factual information? That is, does advice transfer? The key prediction from the framework laid out in Chapter 1 is that it depends on the response to the advice. If the advice is used (in some way), it should lead to (metric) knowledge revision, and thus transfer effects should emerge. In contrast, if the advice is (mostly) rejected, people should retain their metric beliefs, and transfer effects should not be observed.

To address these questions, a modified version of the standard seeding task was used. Participants first generated population estimates for a set of 50 countries, and then they were shown a list with population values for 10 of these countries (information items), before providing a second set of estimates for all 50 countries. The numerical values for the 10 countries always corresponded to the actual populations. However, participants were deceived about the source of this information. They were told that the information was either taken from an almanac (standard seeding) or had been provided by another student who completed the same task and had a certain accuracy level (either 'good' = 95th percentile, or 'average' = 50th percentile).

Besides the stated source credibility of the information, the study included a second manipulation designed to influence the reliance on the provided information. Prior research indicates that people place less weight on advice when it is distant from their own beliefs, compared to when it is close. Therefore, two different sets of information items were used, one consisting of countries for which people's beliefs about national populations were fairly accurate (i.e., low metric error), and one set consisting of countries for which people's beliefs were fairly inaccurate[5]. Given that, on average, the distance between the population values of the list items and people's initial estimates will be greater for the latter

---

[5] This was determined on the basis of a normative study (see Methods section).

than the former list, one might expect that people will be more inclined to use the information from the former list. This is because these values will more likely be consistent with people's initial beliefs and thus might be viewed as confirmation (Nickerson, 1998). In contrast, if people use the list items to extract global metric information (rather than item-level information), then there might be no differences between the information items because the lists were matched in terms of the metric properties of the population values.

2.2 Method

2.2.1 Participants

Two hundred and seventy introductory psychology students participated in the experiment in exchange for partial course credit. All participants were born in Canada and had English as their native language.

2.2.2 Materials

The one hundred most populous countries at the time (Central Intelligence Agency, 2011) were selected as stimulus materials (excluding the four most populous countries [China, India, United States, Indonesia], as well as Canada). The actual populations of these countries ranged from 6 to 203 million. From this set, two subsets of 10 countries were sampled that served as *information items*. Information items are those countries for which participants received feedback or advice regarding the countries' national populations. The selection procedure was as follows: the 100 countries were rank-ordered according to their actual populations and, starting from the top, were divided into 10 blocks of 10 countries. Then, on the basis of normative population estimation data (*N*=124), the country with the greatest and smallest metric error (i.e., order of magnitude error; Brown, 2002) was selected from each block. Thus, two sets of information items were created, one consisting of countries for which people typically have fairly accurate beliefs about national populations (*confirmatory set*), and one consisting of countries for which people typically have fairly inaccurate beliefs about national populations (*disconfirmatory set*). Given this systematic sampling procedure, both sets were representative samples of the reference class, and the actual populations of countries in these two sets did not differ in terms of their mean, $t(18) = -0.04$, $p = 0.97$, or variance, $F(9,9) = 0.79$, $p = 0.72$. The confirmatory and disconfirmatory sets of countries are listed in Appendix A. The numerical population information that was provided for these countries always corresponded to the actual populations rounded to the nearest million.

The remaining 80 countries constituted the pool of *transfer items* (i.e., countries for which no population information was provided). For each participant an individual set of 40 transfer countries was sampled according to the following procedure. The 80 countries were rank-ordered by their true population and divided into 20 blocks of four countries. From each block, two countries were randomly chosen.

2.2.3 Procedure

Participants were tested individually on a computer in a lab-based setting. The experimental session consisted of three tasks: a knowledge rating task, an initial estimation task, and, after the exposure to population information for a subset of items, a final estimation task. During the knowledge rating task, participants were presented with each country and rated on a 10-point rating scale how much they knew about this country in general. This task is commonly used in seeding experiments to examine the role of item familiarity in real-world estimation (Brown & Siegler, 1992, 1993, Brown et al., 2002). In the present study, the task was mainly included to ensure that participants had been exposed to each country before the first estimation task.

During the first estimation task, participants were asked to estimate, to the best of their ability, the current populations of 50 countries. After each response, they also indicated how confident they were that their estimate was accurate using a 1 (*not confident at all*) to 5 (*very confident*) rating scale. The instructions noted that the countries were drawn from the 100 most populated countries in the world. Each trial was initiated by the participant with the spacebar. Then the estimation question appeared in the top half of the screen and participants entered their response into an input field below the question. After confirming their response by pressing the *enter* key, the confidence rating question along with a 5-point rating scale were displayed in the bottom half of the screen. Participants entered their rating and confirmed it with the *enter* key. This caused the program to proceed to the next trial.

After completing the first round of estimations, participants were shown a list of 10 countries names (either the confirmatory or disconfirmatory set) along with population values (in millions) for each country. These values always corresponded to the actual populations of these countries. However, depending on the source credibility condition, participants were led to believe that these values were obtained from different sources. In the seeding ('almanac') condition, participants were informed that the population values were taken from the *World Factbook* and thus corresponded to the actual populations of these countries. In the 'good' advisor condition, participants were told that the numerical values were estimates that had been "provided by an undergraduate student who also completed this estimation task and whose performance was very good (i.e. level of accuracy was within the top 5% of all participants)." In the 'average' advisor condition, the instructions stated that the population values were estimates that had been "provided by an undergraduate student who also completed this estimation task and whose performance was about average (i.e. about half of the participants provided more accurate estimates than this person, and about half provided less accurate estimates)".

Participants were asked to consider this new information carefully. To ensure that every item was processed, participants were prompted with each country name on the list (in a random order) and had to enter the numerical value that was provided[6]. After this familiarization phase, participants were presented with the 50 countries again and were asked to provide a final estimate and confidence rating for each. The procedure was identical to the first estimation task with the exception that the list with the 10 information items was displayed on the left side of the screen throughout this phase. The information items were ordered by their population values (from greatest to smallest). The presentation order of country names was randomized separately for every experimental phase and every participant.

2.3 Results and Discussion

The focus of the following analyses lies on the estimation data. I report separate analyses for the information items and transfer items. These analyses involve the standard measures used in the advice taking and seeding literatures. First, the *weight on self* statistic (WS) is an index of the relative weighting of prior beliefs and new information (also see Table 2.1)[7].

---

[6]In order to ensure that the familiarization task did not lead participants to generate inferences that they were expected to use the new information (Grice, 1975, Schwarz, 1994), a similar experiment was conducted without the familiarization task. The results of this experiment suggest that the familiarization task did not induce Gricean inferences.

[7]Often, a version of this statistic is used where the terms in the numerator and denominator are defined as absolute differences (Bonaccio & Dalal, 2006). This,

WS = (Final Estimate-Information)/(Initial Estimate-Information)

Second, *order of magnitude error* (OME) is a measure of metric accuracy:

$$\text{OME} = |\log_{10}(\text{Estimated Value}/\text{Actual Value})|$$

OME describes the (absolute) discrepancy between the estimated and actual value of a target item in terms of orders of magnitude (Brown, 2002; Nickerson, 1981). For example, an OME of 0 means that the point estimate was on target, and an OME of 1 means that the target was under- or overestimated by one order of magnitude. Third, Spearman rank-order correlations between estimated and actual values serve as a measure of mapping accuracy.

Table 2.1 Mapping of response modes onto weight on self (WS) statistic

Response mode	Values of WS statistic
Overshoot	< -0.05
Adoption	>= -0.05, < 0.05
Assimilation	>= 0.05, < 0.95
Rejection	>= 0.95, < 1.05
Contrast	>= 1.05

Note. Adoption and rejection modes were defined as a range of WS values around 0 and 1 respectively to account for minor deviations due to, for example, rounding.

---

in effect, limits the range of values that the statistic can take to values between 0 and 1. This practice is problematic for an analysis on a response mode level because it converts overshoot and contrast responses into assimilation.

2.3.1 Response Modes and Accuracy

The initial and final estimates for the 10 information items were extracted from each participant's responses. Cases where the initial estimate was equal to the information (actual) value were removed (69 observations, or 2.6% of the data) because these responses could not be unequivocally assigned to one response mode.

Response modes. The distribution of the WS statistic is shown in Figure 2.1. The different response modes are color-coded[8]. Several aspects about these distributions are noteworthy: First, participants' judgments were strongly influenced by the numerical information provided to them. This is reflected, on the aggregate level, in a median WS statistic close to 0 in all conditions. Second, the adoption rate decreased as the stated source credibility declined from the almanac to the average advisor level. That is, as expected, the adoption rate was at ceiling when participants were told that the values were taken from an almanac, but decreased to about 80% when the source was a highly credible student, and reached about 40% for the student framed as average. Third, the adoption rate decreased as a function of source credibility at about the same rate for confirmatory and disconfirmatory sets of information items. Finally, rejection responses were surprisingly rare, even for the lowest source credibility level (about 5%). Thus, with the present paradigm, neither egocentric advice

[8] The absolute frequencies of each response mode are summarized in Figure 4.3.

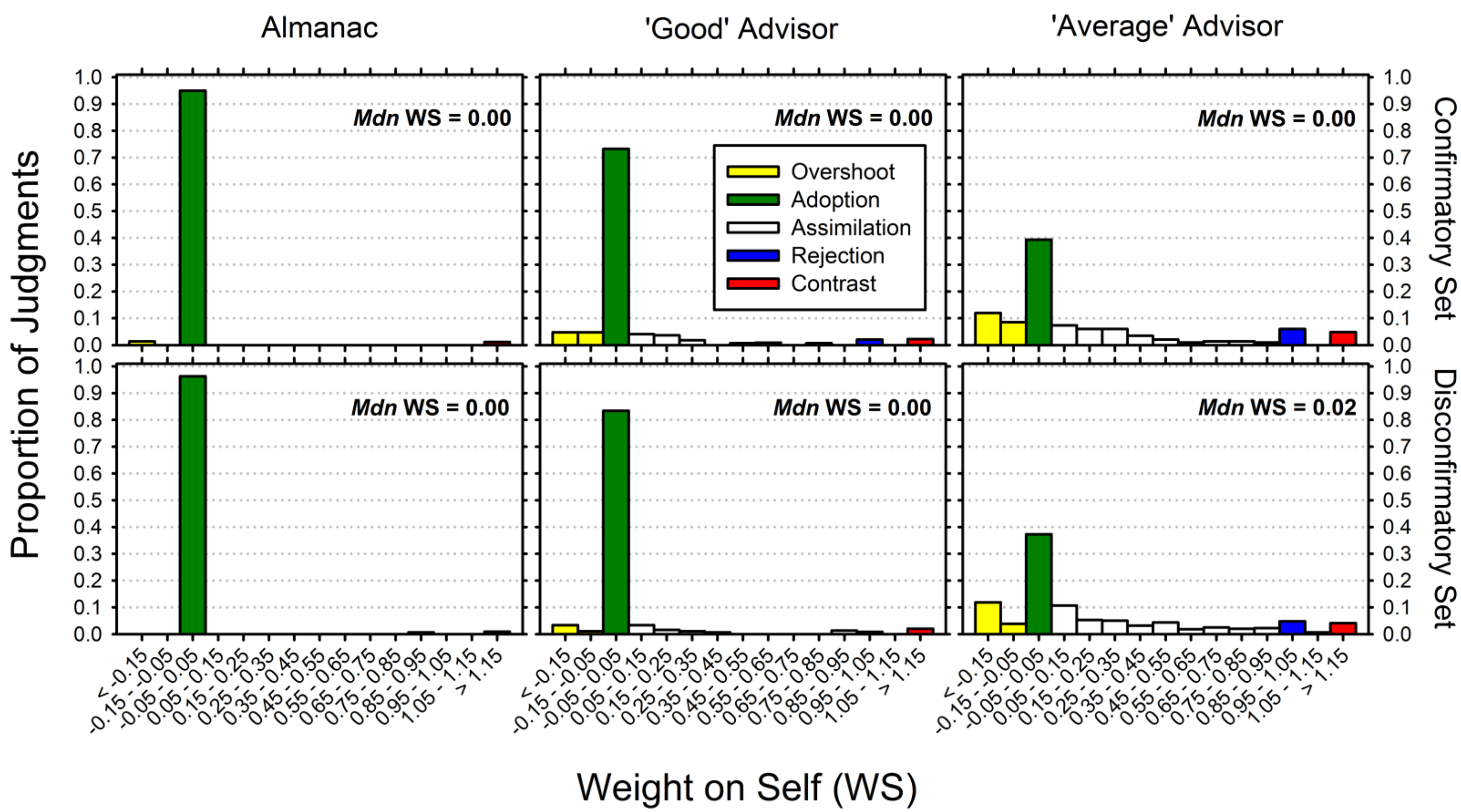

Figure 2.1 Frequency distribution of the binned weight on self (WS) statistic for the information items in Experiment 1. The respective response modes (overshoot, adoption, assimilation, rejection, and contrast) are color-coded. Mdn = Median.

discounting, nor the tri-modal distribution of the WS statistic were replicated (Soll & Larrick, 2009).

Because response modes other than adoption were relatively infrequent, the statistical analysis focused on estimating the odds of adoption (i.e., the probability of adopting the information rather than using any other of the response modes). The data were analyzed using a logistic mixed effects model with subjects and items as random effects. This analysis confirmed pattern evident in Figure 2.1. The model only revealed a significant effect of source credibility, $\chi^2(2) = 146.29$, $p < .001$. Neither the effect of item set, $\chi^2(1) = 0.95$, $p = .33$, nor the interaction was significant, $\chi^2(2) = 2.26$, $p = .32$. The odds of adoption were greater when the information was presented as coming from an almanac rather than a highly credible student (OR = 10.15[9], 95% CI [4.23, 23.83]), and the odds of adoption were also greater with the highly credible student compared to the student of average performance (OR = 22.4, 95% CI [10.34, 48.57]).

Accuracy. Unlike the response mode analysis, the accuracy analyses were conducted on a subject level rather than an individual response level. For each participant the mean OME and rank-order correlation between estimated and actual populations were computed for both the initial and final estimates. Panel A

[9]This odds ratio estimate might be somewhat inflated due to the fact that the adoption rate was at or almost at ceiling in both of these levels of source credibility.

of Figure 2.3 depicts the change in metric accuracy, and Panel B of this figure shows the change in mapping accuracy.

On a metric level, post-information estimates were, on average, more accurate than pre-information estimates. This was true in all conditions. In order to determine if the average metric error of the final estimates differed as a function of source credibility, the data were analyzed with an ANCOVA using OME of the final estimate as the dependent variable, OME of the initial estimate as a covariate, and source credibility as a factor. Because the two item sets

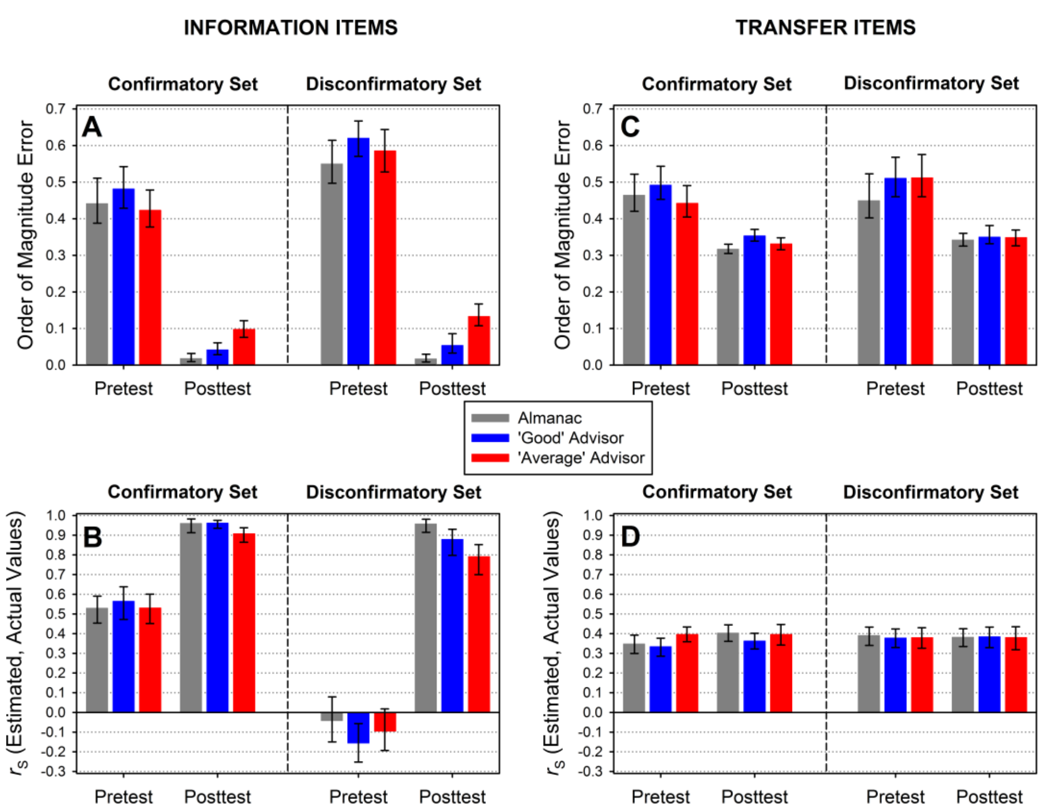

Figure 2.2 Metric and mapping accuracy for the information items (Panels A and B) and the transfer items (Panels C and D). Error bars represent 95% bootstrap confidence intervals of the mean.

differed, by design, in terms of their initial OME, separate analyses were conducted for the confirmatory and disconfirmatory sets. For both sets, the pattern was essentially the same. The metric error of final estimates was slightly greater when the information came from the 'average' rather than the 'good' advisor (contrast estimate = 0.06, 95% CI [0.03, 0.08] for confirmatory set, contrast estimate = 0.08, 95% CI [0.05, 0.11] for disconfirmatory set). The difference between the two highest levels of source credibility ('almanac' vs. 'good') was minimal (contrast estimate = 0.03, 95% CI [0.00, 0.05] for confirmatory set, contrast estimate = 0.03, 95% CI [0.00, 0.07] for disconfirmatory set).

Parallel analyses were conducted for mapping accuracy[10]. The rank-order correlations between estimated and actual population improved for post-information estimates in all conditions. The rank-order correlations of the final estimates were not statistically different between the two highest source credibility levels in either set (contrast estimate = 0.002, 95% CI [-0.04, 0.04] for confirmatory, contrast estimate = -0.06, 95% CI [-0.15, 0.02] for disconfirmatory set). But in both cases, the rank-order correlations for the 'average' advisor were slightly weaker (contrast estimate = -0.05, 95% CI [-0.01, -0.10] for confirmatory set, contrast estimate = -0.10, 95% CI [-0.18, -0.01] for disconfirmatory set).

---

[10] Because for some participants the final rank-order correlations were perfect, an r-to-z transformation was not possible for the mapping accuracy analysis of the information items.

In sum, a lower stated source credibility resulted in slightly worse metric and mapping accuracy of the final estimates. Because the true values were used as information, this pattern reflects the different adoption rates observed in the response mode analysis. However, the absolute accuracy gain was enormous, which again is consistent with the high overall adoption rates.

2.3.2 Seeding Effects and the Transfer of Advice

In this section, I report analyses of the estimates for transfer countries. First, a metric and mapping accuracy analysis is presented and then the relationship between the size of the transfer effect and the initial deviation from the provided information is examined.

Accuracy. The mean by-subject OME and rank-order correlations for the initial and final estimates are shown in Panels C and D of Figure 2.2. In terms of metric accuracy, the average OME for the final estimates decreased relative to the initial judgment task by about the same amount in all conditions. A mixed ANOVA with OME as the dependent variable, judgment (initial vs. final) as a within-subjects factor, and source credibility and item set as between-subjects factors indicated that only the main effect of judgment was reliable, $F(1, 264) = 206.89$, $p < .001$, partial $\eta^2 = .44$. All other effects did not reach significance (all Fs < 2.2, all ps > .11).

On a mapping level, the analogous ANOVA with the r-to-z transformed rank-order correlations as the dependent variable indicated that neither the effect of judgment, $F(1, 264) = 2.60$, $p = .11$, partial $\eta^2 = .01$, nor any of the other effects were significant (all Fs < 2.03, ps > .16). Thus, these data suggest that people draw inductive generalizations from advice in a similar way as they do from factual information. That is, even though there is uncertainty associated with the information regarding its accuracy, people still use it to update their metric knowledge (at least on an aggregate level). In other words, even though the degree of advice utilization for the information items differed, the advice still elicited transfer effects of roughly the same size.

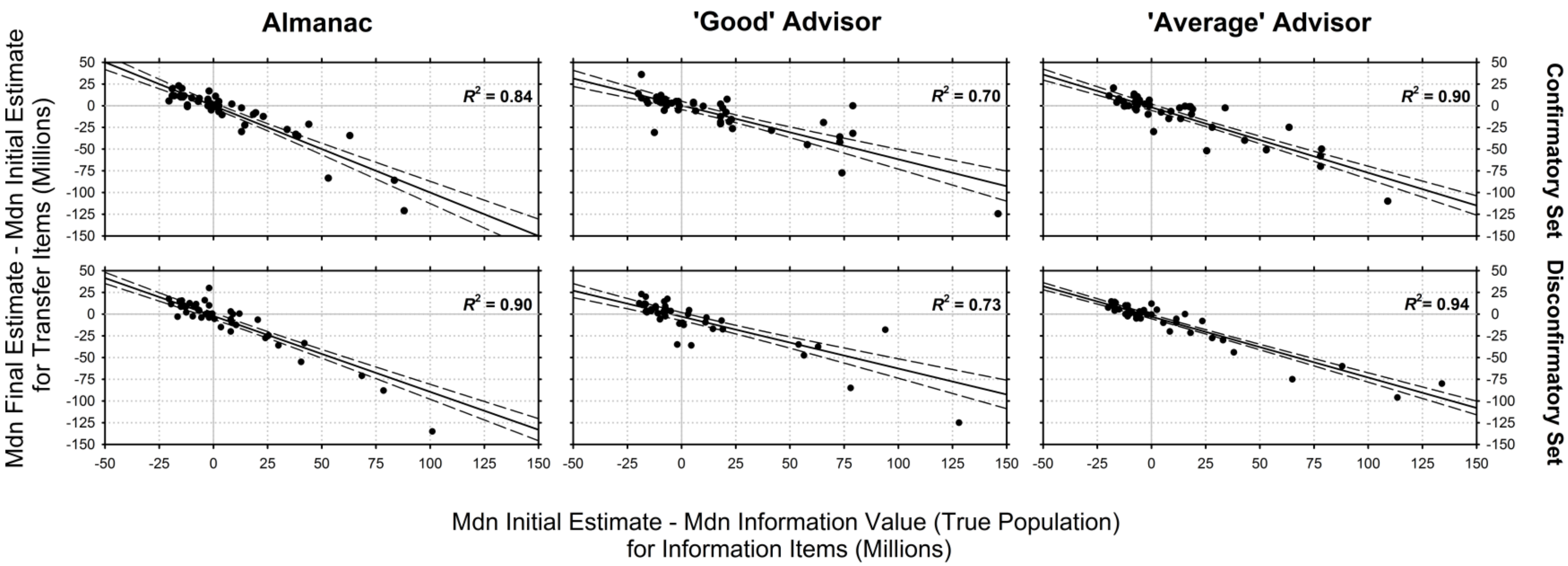

Figure 2.3 Relationship between amount of initial overestimate (underestimate) for the information items and subsequent adjustments of estimates for the transfer items in Experiment 1. Mdn = Median. Regression lines fitted by least-squares regression; bands represent 95% confidence intervals. Mdn = Median.

Individual differences. Why does the aggregate transfer effect not vary as a function of (aggregate) advice utilization? According to the Metrics and Mappings framework, the mechanism underlying the seeding effect is that people update their global metric beliefs about the target dimension. Thus, people with very inaccurate initial metric beliefs should show a greater metric shift than people with relatively accurate initial metric beliefs who might shift, on average, very little or not at all (Brown & Siegler, 1993). Therefore, a large part of the transfer effect might be driven by only a subset of people, people with inaccurate prior metric beliefs. And, if we further assume (a) that these people are more likely to utilize the information in some way (i.e., to not reject it), and (b) that item-level *adoption* of the information (as is typically observed in seeding) is not a necessary condition for updating global beliefs about the metric, then one could explain why, on an aggregate level, transfer is similar irrespective of the credibility of the information.

To examine this account, Figure 2.3 plots, for each participant, the difference between the median final and initial estimate for the transfer items (i.e., the transfer effect) as a function of the difference between the median initial estimate and the median information value (i.e., true value) for the information items. As is evident in the graph, these relationships are linear. That is, if the provided information indicated that the estimates for those items were, on average, too low (high) by a certain amount, then participants tended to adjust their post-information estimates downward (upward) by roughly that same amount. Furthermore, these graphs indicate that for many participants, the transfer effect was quite small, in particular when the difference between the initial estimates and the information was relatively small. Individual paired-samples t-tests comparing initial and final OME indicated that roughly half of the participants in each condition showed a statistically reliable reduction in OME. The proportions ranged from .49 to .59. These participants were also the ones whose initial distance from the information was greater, compared to the ones who did not show the transfer effect (15.5 million vs. 8 million; mean OME: 0.63 vs. 0.42).

Because the rejection rate in this experiment was so low, it was not possible to perform an analysis of the transfer effect as a function of response mode. This will be done in Experiments 2 and 3.

2.4 Conclusions

The present study integrated the seeding and advice-taking paradigms to better understand (a) how source credibility affects information utilization on a response rather than aggregate level, and (b) to determine if and when people draw inductive generalizations from numerical information of uncertain accuracy.

The experiment revealed that source credibility effects are primarily driven by the adoption rate. This supports the view that people have control over how they assess new information and that they use it in different ways. Furthermore, this finding provides additional evidence against the view that people rely on an anchor-and-adjustment process in this task (Lim & O'Connor, 1995). Overall, the rate of advice utilization was higher than is usually observed. This could be due to participants viewing population estimation as a difficult task (Gino, 2008). Another possibility is that the absence of egocentric advice discounting was due to task characteristics, namely the fact that in the current paradigm the initial estimates were not re-presented during the final judgment task. In the standard advice-taking task, both the initial estimate and the advisor estimate are presented. The next experiment explored this possibility.

Furthermore, the present experiment showed that exposure to advice can lead to transfer. The size of these effects was found to be independent of the source credibility associated with the presented information. However, an individual differences analysis indicated that transfer effects are not universal. They are mainly driven by those people who have inaccurate prior beliefs about the target dimension. The next experiment will explore in more detail how the transfer effect is related to the use of the presented information.

3 Experiment 2 - On the Interaction of Task Characteristics and Response Modes in Advice Taking

3.1 Introduction

One of the striking findings of Experiment 1 was the absence of egocentric advice discounting. This is surprising as the effect is typically robust and obtained with comparable materials (Soll & Larrick, 2009). Thus, this experiment was designed to explore why the adoption rate in the hybrid seeding/advice-taking paradigm was comparatively high. Furthermore, if the rejection rate were to increase in the present experiment, this would also allow a test of predictions regarding the relationship between response mode and transfer.

The social comparison literature indicates that people tend to believe that their own skills and abilities are better than those of the average person in domains where absolute skill/ability levels are high—the *above-average effect* (Kruger, 1999). In contrast, in domains where absolute skill level is low, a *below-average effect* is often found where people believe that their skills and abilities are below-average (Moore, 2007). Therefore, one reason why Experiment 1 did not produce egocentric discounting is because people viewed their own knowledge level of national populations as being below-average within their peer group. Experiment 2 therefore included two levels of source credibility. The numerical information was either presented as coming from a student whose performance in the task was average (same as in Experiment 1), or the accuracy level of the supposed student was left unspecified. If people use comparative ability judgments to determine the skill level of the unspecified student, and if these comparative judgments are egocentric (“If I am bad at this task, this other student must be bad at it, too.”), then advice from the ‘unspecified’ student should be discounted more than the advice from the ‘average’ student.

A different account for the lack of egocentric advice discounting in Experiment 1 is that the hybrid seeding/advice-taking task differed in important respects from the standard advice taking task. One of these differences pertains to the *informational context* of the final judgment. That is, in the standard advice taking task, the decision maker is presented with both the advisor’s estimate as well as his or her own initial estimate. In contrast, the paradigm used in the previous study only showed the advice. Given that people use different response modes when they deal with new information, one possibility is that choice of response mode is influenced by task characteristics, such as informational context. There is ample evidence in the judgment and decision making literature that choice of information processing strategies frequently interacts with specific task characteristics (e.g., Hastie & Park, 1986; Hogarth & Einhorn, 1992; Payne, Bettman, & Johnson, 1993), and choice of response modes in the task at hand could similarly be affected by information context. Therefore, in Experiment 2 either only the 10 advice items were displayed during the final judgment, or the

10 advice items and the initial estimate for a given country. If egocentric advice discounting is in part the result of informational context driving the choice of response mode, then egocentric advice discounting should emerge when the initial estimate is present at the time of the final judgment.

3.2 Method

3.2.1 Participants

One hundred eighty undergraduate psychology students participated in this study in exchange for partial course credit. All participants were born in Canada and had English as their native language.

3.2.2 Materials and Procedure

The materials were the same as in Experiment 1, except that only one set of information items was used (disconfirmatory set). The set of transfer countries and their sampling procedure were not changed. In the conditions in which the source credibility of the new information was left unspecified, the population values were introduced as estimates that had been "provided by an undergraduate student who also completed this estimation task".

The procedure was the same as in the prior study, except that during the final judgment task, half of the participants were presented not only with the information items, but also with their initial population estimate for a given country.

3.3 Results and Discussion

3.3.1 Response Modes and Accuracy

The data analysis was conducted analogous to the one reported for the prior experiment. Thirty-nine observations (2.2% of the information item responses) were removed from the data because the initial estimate was equal to the advice and thus these responses could not be assigned to a single response mode.

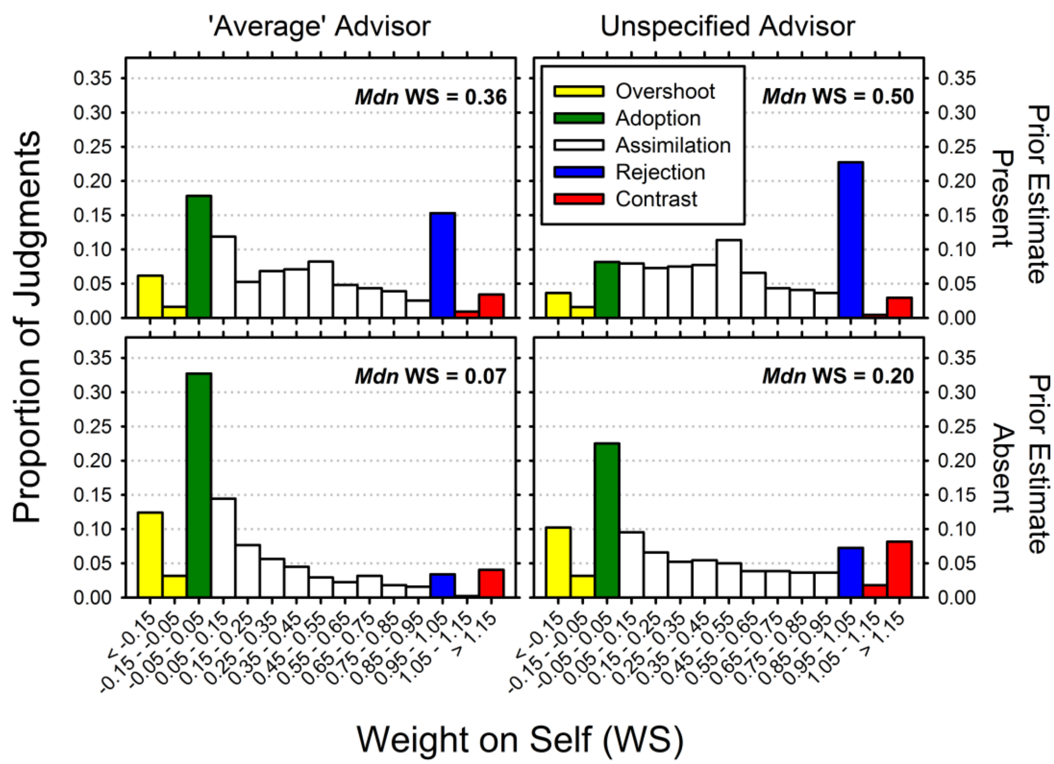

Figure 3.1 Frequency distribution of the binned weight on self (WS) statistic for the information item responses in Experiment 2. The respective response modes (overshoot, adoption, assimilation, rejection, and contrast) are color-coded. Mdn = median.

Response modes. Figure 3.1 depicts the distributions of the WS statistic for each condition. The median WS was never greater than 0.5, indicating that people typically underweighted their own estimate relative to the advisor's estimate. Furthermore, participants weighted their own beliefs more heavily when the credibility of the source was left unspecified, compared to when it was stated that the information came from a source with 'average' accuracy. Overall, the difference in median WS between these two levels of source credibility was about .14. This finding is consistent with the social comparison account outlined above (Kruger, 1999). It suggests that people viewed the accuracy of their own beliefs about national populations as being below-average, and viewed the accuracy of the beliefs of an unspecified peer as being below-average, too. More strikingly, the median WS statistic also indicates participants placed more weight on their own estimate when it was displayed along with the information during the final judgment task. The presentation of the prior estimate was associated with an increase in median WS from 0.07 to 0.36 in the 'average' advisor condition, and an increase from .20 to .50 in the unspecified advisor condition. Furthermore, when both the advice and the prior estimate were presented, the WS distributions approximate the tri-modal shape—with modes at 0, 0.5, and 1—that was found in previous advice taking studies using a more heterogeneous set of questions (Soll & Larrick, 2009). In contrast, when the prior estimate was not presented, the resulting WS distributions are much more skewed towards low WS values. These findings suggest that task characteristics, such as the accessibility of different pieces of information, can have an impact on if and how new information is used.

To examine how source credibility and information accessibility affected the relative use of response modes, a multinomial logistic GEE[11] regression model was fitted to the data (Touloumis, 2015; Touloumis, Agresti, Kateri, 2013). Response mode was treated as a nominal response category. To reduce the number of comparisons, overshoot and contrast responses were removed from the data[12]. The final model included the predictors: source (average vs. unspecified) and prior estimate (present vs. absent), and OME of initial estimates (to take the discrepancy between initial estimate and advice into account). All interaction terms were non-significant (all $|z|$s < 0.96) and were therefore dropped from the model. The model coefficients are reported in Appendix B.

---

[11] GEE = Generalized Estimating Equations. This approach was chosen because each participant in the present study provided multiple responses, and GEE models, in particular when combined with robust variance estimates, allow to account for clustering in the data (Hardin & Hilbe, 2013).

[12] It is common practice in the advice taking literature to either (a) exclude overshoot and contrast responses, (b) to set overshoot values to 0 and contrast values to 1, thus treating these as adoption and rejection, respectively, or (c) to use an alternative version of the WS statistic that involves absolute differences in numerator and denominator, which, in effect, shifts overshoot and contrast responses into the assimilation range.

Table 3.1 Odds ratio (OR) estimates and 95% confidence intervals for the predictors of interest from the multinomial GEE model fitted to the data (information items) in Experiment 2.

Variable	Rejection vs. Adoption		Assimilation vs. Adoption		Rejection vs. Assimilation	
	OR	95% CI	OR	95% CI	OR	95% CI
Source (Average vs. Unspecified)	0.37^{**}	[0.20,0.68]	0.56^{*}	[0.34,0.90]	0.66^{+}	[0.42,1.02]
Prior estimate (Present vs. Absent)	6.54^{***}	[3.43,12.47]	2.45^{***}	[1.50,4.02]	2.67^{***}	[1.67,4.26]

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$; $^{+}p < .10$

The effect size estimates (stated as odds ratios) are shown in Table 3.1. First, comparing rejection vs. adoption, it is evident that both source credibility and presence of prior estimate impacted the relative frequency with which these response modes were used. Specifically, the odds that the information was rejected rather than adopted (conditional[13] on not assimilating) decreased when the information was presented as coming from an average advisor rather than an unspecified advisor (OR = 0.37). Furthermore, the odds of rejecting rather than adopting the information was about 6.5 times greater when the prior estimate was presented during the final judgment task, compared to when it was not presented.

Second, the comparison between assimilation and adoption responses shows that assimilation was slightly less common when the information was presented as coming from an average rather than an unspecified advisor (OR =

---

[13]The estimated odds described in this section are all conditional. However, for the sake of reducing redundancy, this is not explicitly stated for every comparison.

0.56). On the other hand, the odds of assimilation increased when both the advice and the prior estimate were presented during the final judgment task, compared to when only the advice was shown (OR = 2.45).

Finally, the comparison between rejection and assimilation responses indicates that the odds that the information was rejected decreased slightly when the source was more credible (OR = 0.66, effect is marginally significant). Furthermore, the odds of rejection (rather than assimilation) when the prior estimate was present were about 2.7 times the odds of rejection when the prior estimate was absent.

In sum, these analyses suggest that an increase in stated source credibility promotes adoption at the expense of rejection and assimilation responses. Conversely, the presentation of the prior estimate along with the advice strongly promotes rejection and also assimilation, and lessens the frequency of adoption responses (also see Figure 4.3).

Accuracy. The mean OMEs and rank-order correlations for the initial and final estimates are reported in Table 3.2. Exposure to the numerical advice improved metric accuracy in all conditions. An ANCOVA with final OME as the dependent variable, source credibility and presence of prior estimate as factors, and initial OME as a covariate, indicated that the accuracy of final estimates improved more when the information was presented as coming from an 'average' advisor compared to an unspecified advisor, $F(1, 175) = 14.53$, $p < .001$, partial $\eta^2 = .08$. Furthermore, the average accuracy of the final estimates was better when

Table 3.2 Order of magnitude error (OME) and rank-order correlations between estimated and actual population for the information items in Experiment 2

	Prior estimate present				Prior estimate absent			
	Initial estimate		Final estimate		Initial estimate		Final estimate	
Measure	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
OME								
'Average' advisor	0.56	[0.51, 0.63]	0.24	[0.20, 0.28]	0.60	[0.54, 0.66]	0.16	[0.13, 0.20]
Unspecified advisor	0.56	[0.51, 0.61]	0.30	[0.26, 0.34]	0.62	[0.57, 0.68]	0.26	[0.21, 0.32]
Rank-order correlation								
'Average' advisor	-0.08	[-0.16, 0.04]	0.61	[0.50, 0.71]	-0.15	[-0.25, -0.03]	0.78	[0.71, 0.84]
Unspecified advisor	-0.07	[-0.17, 0.04]	0.46	[0.34, 0.56]	-0.14	[-0.24, -0.05]	0.45	[0.30, 0.58]

Note. CI = 95% bootstrap confidence interval of the mean.

the initial estimate was not presented during the advice-taking phase, $F(1, 175) = 10.99$, $p = .001$, partial $\eta^2 = .06$. The interaction was not significant, $F(1,175) = 0.64$, $p = .42$, partial η^2 .004.

Similarly, mapping accuracy improved in all four conditions. The analogous ANCOVA with rank-order correlations indicated that the final rank-order correlations were greater in the average than the unspecified advisor condition, $F(1, 175) = 20.02$, $p < .001$, partial $\eta^2 = .10$, and were also greater, on average, when the initial estimate was absent during the final judgment task, $F(1, 175) = 3.84$, $p = .05$, partial $\eta^2 = .02$. However, the interaction term approached significance, $F(1, 175) = 3.21$, $p = .08$, partial $\eta^2 = .02$, reflecting that the effect of prior estimate was primarily driven by the average advisor condition.

In sum, the results of the accuracy analysis are highly consistent with the observed response mode differences. Given that the actual population values were used as advice, combined with the finding that heightened (perceived) source credibility promotes adoption and that the presentation of the prior estimate promotes rejection, it follows that the average final estimation error should be lower in those contexts where source credibility is relatively high and only the advice is present.

3.3.2 Transfer of Advice

Accuracy. Replicating the transfer effect from Experiment 1, population estimates for transfer countries were, on average, more accurate after participants

Table 3.3 Order of magnitude error (OME) and rank-order correlations between estimated and actual population for the transfer items in Experiment 2

	Prior estimate present				Prior estimate absent			
	Initial estimate		Final estimate		Initial estimate		Final estimate	
Measure	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
OME								
'Average' advisor	0.46	[0.41, 0.53]	0.36	[0.34, 0.39]	0.48	[0.44, 0.54]	0.35	[0.33, 0.38]
Unspecified advisor	0.45	[0.42, 0.51]	0.38	[0.36, 0.41]	0.52	[0.47, 0.58]	0.37	[0.35, 0.39]
Rank-order correlation								
'Average' advisor	0.39	[0.33, 0.44]	0.43	[0.36, 0.48]	0.34	[0.30, 0.38]	0.40	[0.36, 0.44]
Unspecified advisor	0.38	[0.33, 0.43]	0.44	[0.39, 0.48]	0.40	[0.35, 0.44]	0.41	[0.37, 0.45]

Note. CI = 95% bootstrap confidence interval of the mean.

had received advice on the information items (see Table 3.3). On the metric level, a mixed ANOVA with OME as the dependent variable, judgment (initial vs. final) as the within-subjects factor, and source and presentation of prior estimate as between-subjects factors revealed a significant effect of judgment, $F(1, 176) = 120.05$, $p < .001$, partial $\eta^2 = .41$. However, this effect was slightly greater when the prior estimate was not shown (judgment × prior estimate interaction, $F[1, 176] = 8.25$, $p = .005$, partial $\eta^2 = .05$), which is mainly due to somewhat higher initial OMEs in these conditions. The mean OMEs for the final estimates are very similar across conditions. All other effects were non-significant (Fs < 1.46).

On a mapping level, an analogous ANOVA with the r-to-z transformed rank-order correlations indicted that mapping accuracy for transfer countries also slightly improved, $F(1, 176) = 24.47$, $p < .001$, partial $\eta^2 = .12$. However, the improvement was very small (mean pretest $r_S = .38$, mean posttest $r_S = .42$). All other effects were not significant (Fs < 2.46, ps > .12). Therefore, overall these results are consistent with the findings from Experiment 1 and indicate that advice transfers. Again, the aggregate transfer effect was roughly of the same size in all conditions.

Individual differences. Because response modes other than adoption occurred more frequently in this experiment, it was possible to examine the relationship between response mode use and the size of the transfer effect. Participants were categorized into one of four types: if a participant rejected 50% or more of the advice for the information items, the participant was termed a *frequent rejector* (7.2% of participants). The analogous criterion was used for adoption (17.2% were categorized as *frequent adopters*), and assimilation (56.1% were categorized as *frequent assimilators*). Subjects without a response mode preference were assigned to the last category, the *mixed responders* (19.4% of participants)[14].

In Figure 3.2 the size of the transfer effect—the difference between the mean OME of the final and initial estimates for transfer countries—is plotted as a function of the discrepancy between the initial estimates and the advice for the information items (i.e., mean initial OME for the information items). The graph includes a separate plot for each of the four responder types defined above. In each plot, non-parametric regression lines (LOWESS[15]) were superimposed to depict the relationship between the two variables. These plots illustrate two important points. First, for the frequent adopters, frequent assimilators, and mixed responders, the size of the transfer effect increased as the discrepancy between the subject's initial estimates and the advice increased. This pattern is consistent with the results of Experiment 1 and supports the view that people use feedback to revise their metric beliefs about the target dimension (Brown & Siegler, 1993). These data also indicate that *adoption* of all the values (as is the case in seeding) is not a necessary condition for transfer.

---

[14]One participant could have been assigned to either the frequent rejector or frequent adopter category, and was randomly assigned to the former.

[15]LOWESS = locally weighted scatterplot smoothing

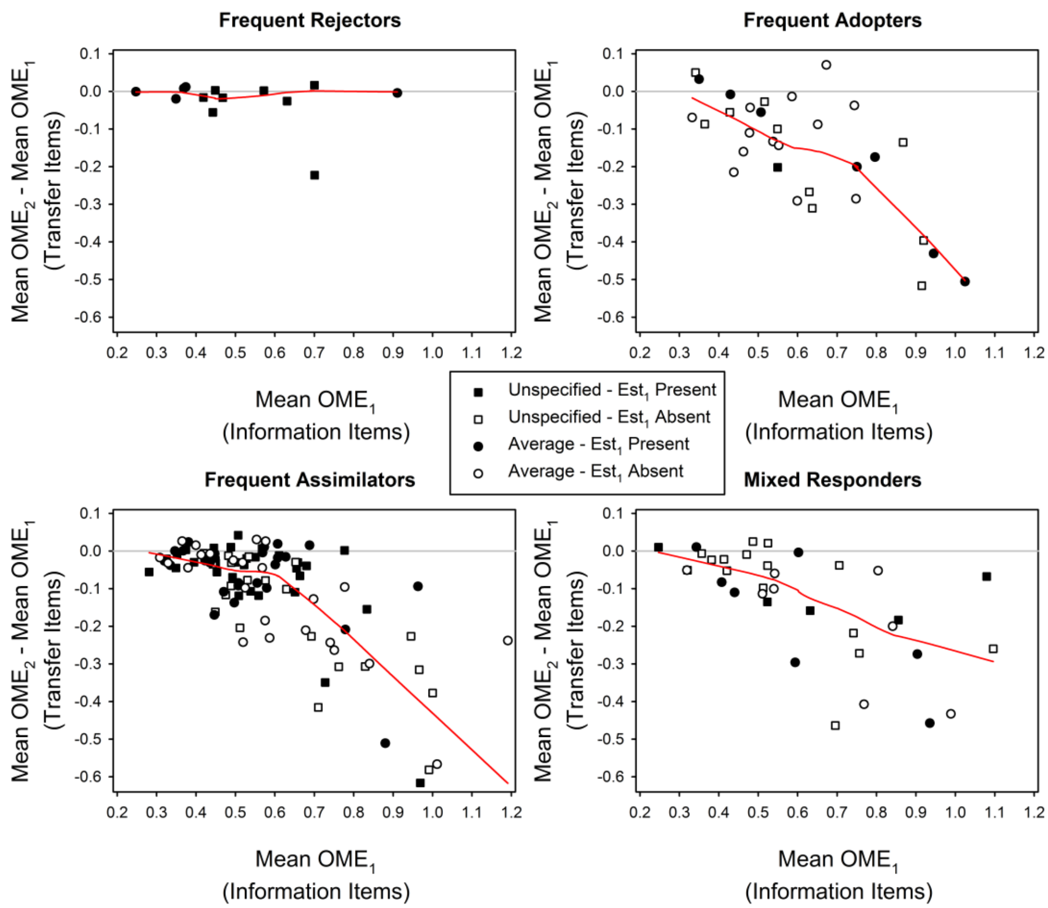

Figure 3.2 Size of the transfer effect as a function of initial discrepancy between estimate and advice (as measured by order of magnitude error) and response mode preference (Experiment 2). Regression lines fitted by locally weighted regression (LOWESS). Est = Estimate; OME = order of magnitude error.

Second, the frequent rejectors provide a striking exception to the pattern observed for the other types of responders. Participants who rejected the advice frequently did not show a transfer effect, and this was true irrespective of the participant's initial OME. These two findings are consistent with the view that knowledge revision involves controlled processing as both the presence and the size of the transfer effect seem to depend on how the new information is initially assessed (as relevant or irrelevant), and how discrepant the new information is from a person's prior beliefs. In other words, these results support strong evidence against the view that exposure to numerical information will inevitably "contaminate" judgment and/or the underlying knowledge base.

3.4 Conclusions

The results of this experiment provided two main insights. First, informational context—defined as the numerical information made accessible to the decision maker during the target judgment—seems to influence, in part, how people respond to new numerical information. In particular, when the prior estimate is present, people are more likely to reject the new information.

Second, this experiment tested a critical prediction derived from the framework outlined in Chapter 1. Namely that knowledge revision is dependent on actively engaging with and using the new information. In contrast, if the information is rejected, no updating occurs (as reflected in the absence of transfer).

4 Experiment 3 – A Paradigmatic Examination of the Role of Informational Context in Advice Taking

4.1 Introduction

The results of the previous experiment indicated that informational context—defined as the numerical information made accessible to the decision maker during the target judgment—had an influence on how people utilized the numerical advice offered to them. Specifically, people tended to reject advice more often when their prior estimate was presented along with the advice values. In contrast, people tended to adopt advice more frequently when only the advice values were shown. This pattern might reflect a preference for immediately accessible information. When only the advice is presented, people are inclined to use it more often because it is directly accessible (and potentially plausible) numerical information. However, if both the prior estimate and the advice are made accessible, people might be egocentrically biased towards their own estimate.

In order to further explore if and how the choice of response modes is driven by the informational context, an experiment was designed that tested all four possible combinations of presenting or not presenting the advice values and the initial estimate, respectively. If people's choice of response mode is partly affected by what information is displayed during the target judgment, then the rejection rate should be greatest when only the prior estimate is shown. Conversely, the adoption rate should be greatest when only the advice is shown.

4.2 Method

4.2.1 Participants

Two hundred twenty-four introductory psychology students participated in this experiment. All participants were born in Canada and had English as their native language.

4.2.2 Materials and Procedure

The materials and procedure were the same as in the previous experiment. When participants were presented with the advice (true values) for the 10 countries, the credibility of the source always remained unspecified. Depending on the condition, participants were also informed that during the final judgment task, either both the list with the advice and their prior estimate for the target country would be displayed on a given trial, only the former or only the latter would be displayed, or neither of the two would be displayed.

4.3 Results and Discussion

The analysis section follows the structure of the previous two experiments.

4.3.1 Response Modes and Accuracy

As was done in the previous studies, any responses where the initial estimate equaled the advice value were removed from the information item data (36 responses, or 1.6% of the data).

Response modes. The WS distributions obtained in each condition are depicted in Figure 4.1. Overall, the manipulation of information accessibility during the target judgment proved highly effective. The median WS ranged from 0.75—when the judgment context only involved the display of the initial estimate —to 0.23—when the judgment context only involved the display of the advice. When one simply interprets these aggregate statistics (a practice that most prior advice taking research has relied on), one would conclude that in the former case participants put more weight on their own opinion than the advisor's (i.e., displayed egocentric advice discounting), whereas in the latter case one would conclude that participants put more weight on the advisor's opinion than their own (i.e., were more easily influenced by new information). That is, depending on the judgment context, one could potentially arrive at diametrically different interpretations of how people deal with new information; even when the characteristics of the source and message, and most of the participant's knowledge state are held constant (as in the present experiment).

As seen before, the WS distributions were multi-modal. This suggests that people responded differently to the advice (see Table 4.1). In order to determine the extent to which contextual factors (i.e., information accessibility) influenced the choice of response modes, the data were analyzed with logistic mixed models to estimate the odds of using a particular response mode rather than any of the other modes. All models included subjects and items as random effects and informational context[16] and initial OME as fixed effects.

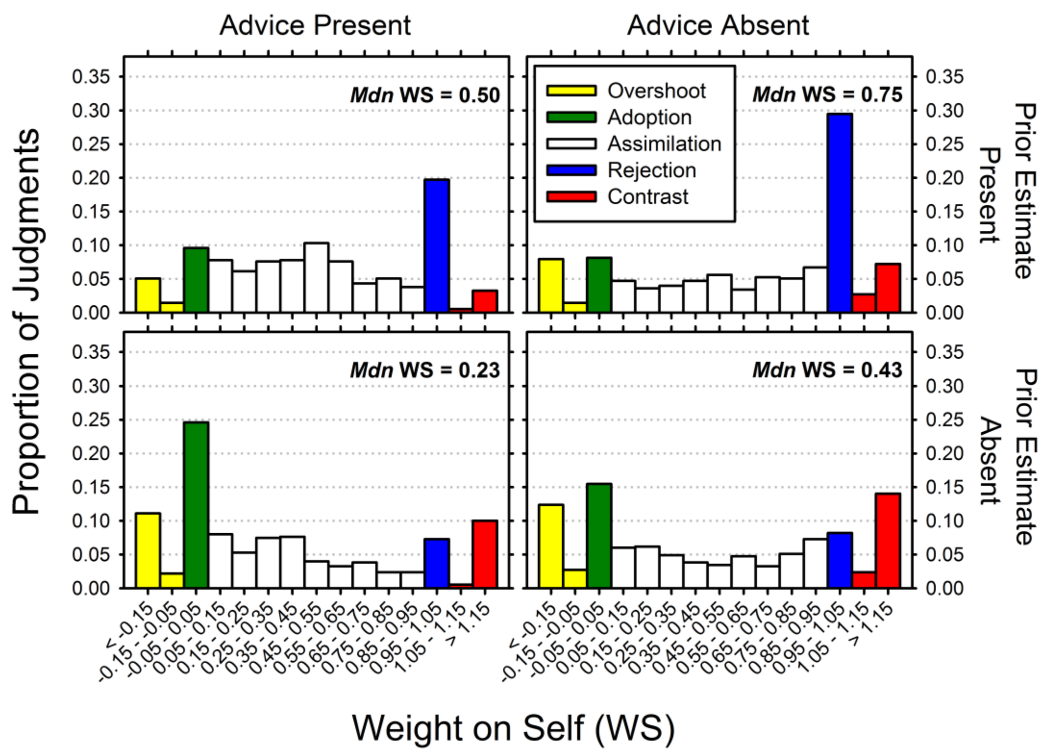

Figure 4.1 Frequency distribution of the binned weight on self (WS) statistic for the information items in Experiment 3. The respective response modes (overshoot, adoption, assimilation, rejection, and contrast) are color-coded. Mdn = Median.

---

[16]Because individual comparisons between the four informational contexts were of primary interest, a single predictor with 2 (Advice: present vs. absent) × 2 (Prior estimate: present vs. absent) levels was included in the model.

Table 4.1 Frequency of response modes (in percent) as a function of which information was displayed during the final judgment task (Experiment 3)

Displayed information	Adoption	Rejection	Assimilation	Overshoot	Contrast
Advice only	24.6	7.3	44.3	13.3	10.6
None	15.5	8.2	44.8	15.1	16.4
Advice + Est_1	9.6	19.7	60.4	6.5	3.8
Est_1 only	8.1	29.5	43.0	9.4	10.0

Note. Est_1 = initial estimate.

The experimental conditions in Table 4.1 are rank-ordered by the adoption rate. This organization reveals one of the main patterns in the data, namely the inverse relationship between the frequency of adoption and rejection in the different informational contexts. Specifically, the odds of adoption were smallest whenever the initial estimate was displayed (i.e., the *Est_1 only* and *Advice + Est_1* conditions; the adoption rate in these two conditions did not differ from one another, OR = 1.06, 95% CI [0.40, 2.82], $z = 0.11$, $p = .91$). The odds of adoption increased when neither the initial estimate nor the advice were displayed (relative to when both were displayed), OR = 2.64, 95% CI [1.03, 6.76], $z = 2.03$, $p = .04$. Finally, the odds of adoption were even greater when only the advice was present (compared to when neither was present), OR = 2.51, 95% CI [1.05, 5.96], $z = 2.08$, $p = .04$.

Conversely, the odds of rejection were smallest whenever the prior estimate was not displayed (i.e., the *Advice only* and *None* conditions; here, it made virtually no difference whether or not the advice was present, OR = 0.84, 95% CI

[0.46, 1.52], $z = -0.59$, $p = .56$). However, the odds of rejection increased when both pieces of information were present (compared to none being present), OR = 2.92, 95% CI [1.70, 5.01], $z = 3.90$, $p < .001$, and even further increased when only the initial estimate was displayed, OR = 1.90, 95% CI [1.18, 3.04], $z = 2.66$, $p = .008$.

The pattern for assimilation responses looks quite different from the ones for adoption and rejection. When both the advice and the prior estimate were present, the odds of assimilation were greater than in any of the other informational contexts; that is, relative to the advice only context, OR = 2.33, 95% CI [1.50, 3.62], $z = 3.74$, $p < .001$), when neither the advice nor the initial estimate were present, OR = 2.23, 95% CI [1.43, 3.46], $z = 3.56$, $p < .001$, or when only the initial estimate was displayed, OR = 2.09, 95% CI [1.34, 3.25], $z = 3.26$, $p = .001$.

Overshoot responses were least frequent whenever the initial estimate was present (i.e., the *Est_1 only* and *Advice + Est_1* conditions; the assimilation rate in these two conditions did not differ from one another, OR = 1.72, 95% CI [0.90, 3.26], $z = 1.65$, $p = .10$. Relative to when only the initial estimate was presented, the odds of overshoot increased when only the advice was shown, OR = 2.43, 95% CI [1.30, 4.54], $z = 2.79$, $p = .005$), and when neither advice nor initial estimate was shown, OR = 3.69, 95% CI [1.98, 6.90], $z = 4.10$, $p < .001$.

Finally, the odds that participants’ final estimates moved away from both the advice and the initial estimate (contrast) was smallest when both pieces of information were displayed, but increased when either only the initial estimate, OR = 2.98, 95% CI [1.68, 5.32], $z = 3.72$, $p < .001$, or only the advice were presented, OR = 2.92, 95% CI [1.65, 5.18], $z = 3.67$, $p < .001$. The odds of contrast were greatest when no information was displayed during the judgment, OR (relative to advice only) = 2.04, 95% CI [1.32, 3.16], $z = 3.23$, $p = .001$.

In summary, the response mode analysis revealed the following main patterns: (1) the frequency of adoption and rejection responses had an inverse relationship, (2) assimilation responses were more frequent when both the advice and the prior estimate were present, and (3) overshoot and contrast were less frequent when both advice and prior estimate were present. The implications of these findings will be discussed in Section 4.5.

Accuracy. In all conditions, participants' post-advice estimates improved both in terms of metric accuracy as well as mapping accuracy (see Table 4.2). An ANCOVA with mean OME of the final estimate as the dependent variable, initial mean OME as a covariate, and presence of advice (present vs. absent) and presence of the prior estimate (present vs. absent) as between-subjects factors indicated that the metric accuracy of the final judgment was better when the advice was present rather than absent, $F(1, 219) = 33.35$, $p < .001$, partial $\eta^2 = .13$, and was worse when the initial estimate was present rather than absent, $F(1, 219) = 8.65$, $p = .004$, partial $\eta^2 = .04$. The interaction was not significant, $F(1, 219) = 2.47$, $p = .12$, partial $\eta^2 = .01$.

Table 4.2 Order of magnitude error (OME) and rank-order correlations between estimated and actual population for the information items in Experiment 3

	Prior estimate present				Prior estimate absent			
	Initial estimate		Final estimate		Initial estimate		Final estimate	
Measure	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
OME								
Advice present	0.61	[0.56, 0.66]	0.29	[0.25, 0.34]	0.55	[0.51, 0.60]	0.25	[0.21, 0.28]
Advice absent	0.59	[0.55, 0.66]	0.43	[0.39, 0.48]	0.63	[0.57, 0.70]	0.35	[0.32, 0.39]
Rank-order correlation								
Advice present	0.00	[-0.09, 0.09]	0.59	[0.48, 0.67]	-0.06	[-0.15, 0.04]	0.56	[0.46, 0.29]
Advice absent	-0.04	[-0.11, 0.06]	0.21	[0.12, 0.64]	-0.09	[-0.18, 0.01]	0.28	[0.18, 0.38]

Note. CI = 95% bootstrap confidence interval of the mean.

The analogous ANCOVA with the rank-order correlations revealed that in the presence of the 10 advice items, participant's rank-order correlations of final estimates were better compared to when the advice items were absent, $F(1, 219) = 49.58$, $p < .001$, partial $\eta^2 = .19$. The presentation of the prior estimate did not affect rank-order correlations, $F(1, 219) = 0.61$, $p = .44$, partial $\eta^2 = .003$. The interaction was also not significant, $F(1, 219) = 1.20$, $p = .28$, partial $\eta^2 = .005$.

4.3.2 Transfer of Advice

Accuracy. Consistent with the results of the previous experiments, the exposure to advice for a subset of countries affected the estimates for transfer countries (see Table 4.3). Specifically, results of a mixed ANOVA indicated the mean OME of post-advice estimates was smaller than for initial estimates, $F(1, 220) = 99.08$, $p < .001$, partial $\eta^2 = .31$. The transfer effect trended towards being slightly smaller when the prior estimate was present rather than absent, $F(1, 220) = 3.04$, $p = .08$, partial $\eta^2 = .01$ for the judgment (initial vs. final) × presence of prior estimate (present vs. absent) interaction. All other effects in the mixed ANOVA were non-significant (Fs < 2.17, ps > .14).

For the (r-to-z transformed) rank-order correlations, a mixed ANOVA revealed a very small, but significant, improvement in the rank-order correlations for the post-advice estimates (from mean $r_S = .38$ to mean $r_S = .42$), $F(1, 220) = 20.48$, $p < .001$, partial $\eta^2 = .09$. All other effects were not significant (Fs < 1).

Table 4.3 Order of magnitude error (OME) and rank-order correlations between estimated and actual population for the transfer items in Experiment 3

	Prior estimate present				Prior estimate absent			
	Initial estimate		Final estimate		Initial estimate		Final estimate	
Measure	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
OME								
Advice present	0.52	[0.47, 0.58]	0.42	[0.38, 0.46]	0.47	[0.43, 0.53]	0.37	[0.34, 0.41]
Advice absent	0.52	[0.47, 0.59]	0.43	[0.40, 0.49]	0.54	[0.48, 0.61]	0.39	[0.36, 0.42]
Rank-order correlation								
Advice present	0.39	[0.34, 0.43]	0.42	[0.38, 0.46]	0.38	[0.33, 0.42]	0.42	[0.37, 0.46]
Advice absent	0.39	[0.35, 0.43]	0.43	[0.39, 0.46]	0.36	[0.32, 0.41]	0.42	[0.38, 0.45]

Note. CI = 95% bootstrap confidence interval of the mean.

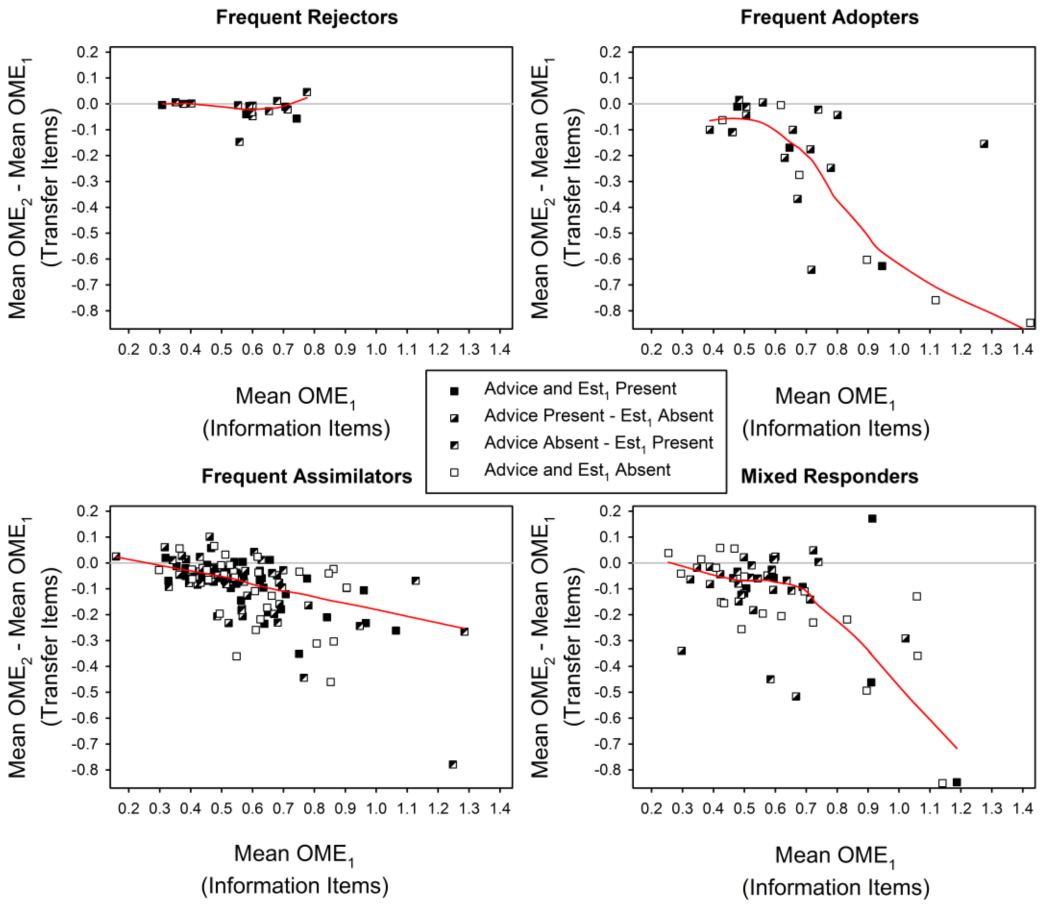

Figure 4.2 Size of the transfer effect as a function of initial discrepancy between estimate and advice (as measured by order of magnitude error) and response mode preference (Experiment 3). Regression lines fitted by locally weighted regression (LOWESS). Est = Estimate; OME = order of magnitude error.

Individual differences. The same categorization scheme was used as in the previous experiment. Participants were classified[17] as frequent rejectors (7.6%), frequent adopters (11.1%), frequent assimilators (55.4), and mixed responders (25.9%). Figure 4.2 shows, for each responder type, the size of the transfer effect as a function of the average absolute discrepancy between initial estimates and the advice values (i.e., mean OME). As in the previous study, the size of the transfer effect increased as a function of initial OME for all responder types, except for the frequent rejectors. The frequent rejectors showed virtually no transfer effect, irrespective of initial OME. Again, this suggests that people have to use the advice (in some form) in order to trigger processes of knowledge revision.

4.4 Conclusions

The results of the present experiment suggest that the informational context in which the target judgment is made plays an important role in how people respond to numerical advice. In particular, adoption and rejection are affected by whether the advice or the prior estimate is displayed during the target judgment. Why might this be? One possibility is that it is related to the relatively low level of knowledge of country populations. Both the comparatively high rate of advice utilization and the relatively high initial metric error, suggests that this

[17]One participant qualified for two of the categories and was randomly assigned to one of the two.

was a difficult task for the subject population. Thus, for several target countries, people's subjective range of plausible population values might have been fairly wide, such that occasionally both the advice value and the prior estimate fell into this range. In these cases of metric indifference, participants then preferred to retain their initial estimate when it was displayed, but adopted the advice when their initial estimate was not displayed (and occasionally opted to combine values when both were displayed). From this perspective, the effect of informational context should be significantly reduced if target item were used that are taken from a better calibrated knowledge domain.

4.5 Response Modes: Summary

Figure 4.3 summarizes the relative frequency with which each response mode was used in Experiments 1-3. First, the graph for adoption response shows that the adoption rate incrementally decreased as the stated source credibility of the numerical information decreased. This pattern is important because it shows that source credibility effects are driven by the adoption rate (rather than an average incremental increase in assimilation). Furthermore, it again illustrates that people have control over whether to adopt or not to adopt new information.

Second, the source credibility and informational context manipulations drove the rejection rate from 0% to about 30%. This finding demonstrates that people can deal with numerical information without being influenced by it.

Third, the assimilation rate was generally greater when both the prior estimate and the numerical advice were displayed during the final judgment task. Given that averaging is the best aggregation strategy in many judgment environment (Soll & Larrick, 2009), one implication of the present finding is that in order to promote averaging, one should ensure that both initial estimate and advice are made accessible to the decision maker.

Finally, overshoot and contrast responses occurred occasionally, in particular when the source credibility was low. The overall rate of these response modes was somewhat greater than is typically observed in standard advice taking studies. This might be related to the fact that in the present experiments people were exposed to a sample of 10 information items and thus might have revised their beliefs not on an item-level, but on a more global (metric) level.

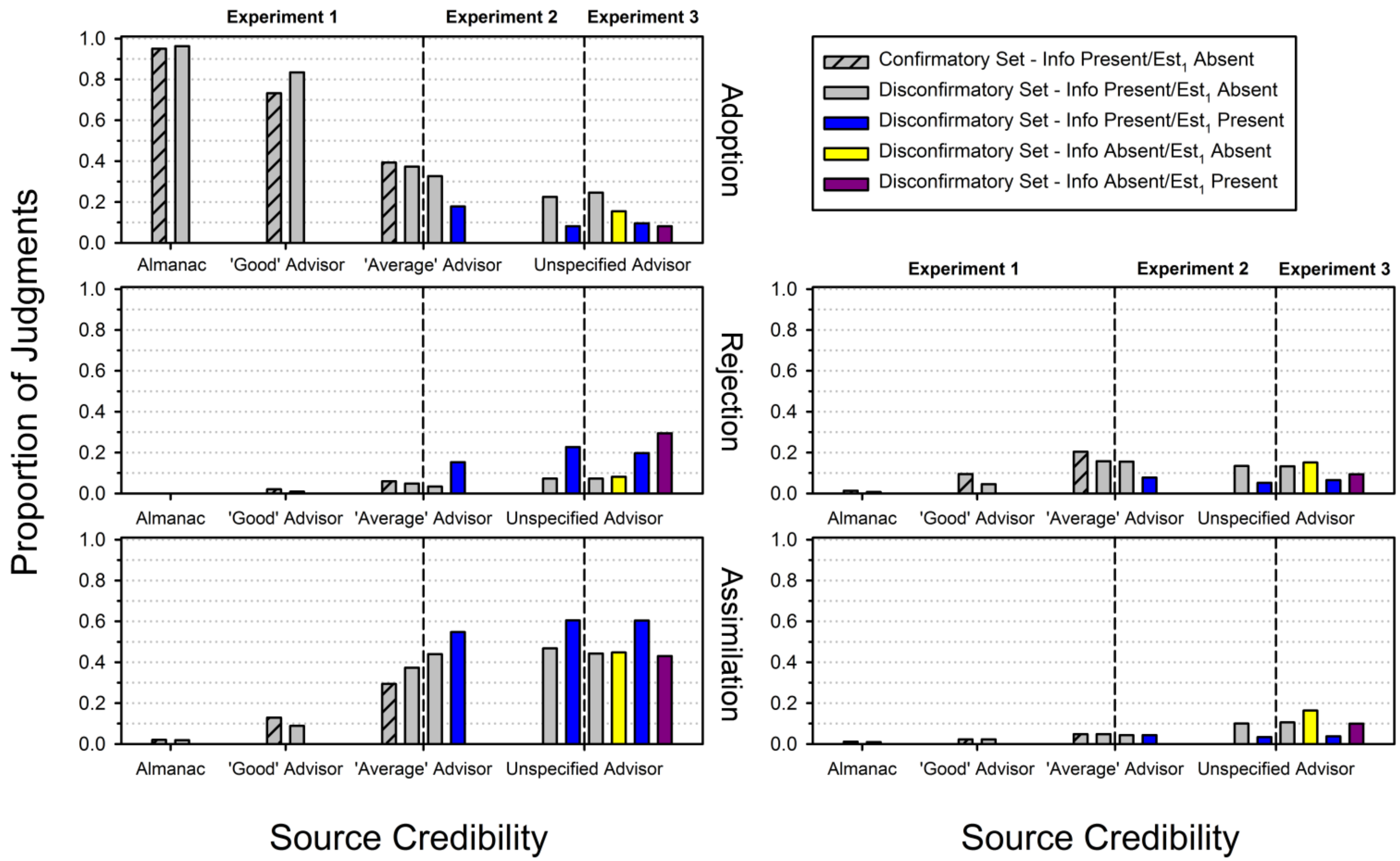

Figure 4.3 Relative frequency of response modes as a function of source credibility and informational context in Experiments 1-3. Info = Information; Est = Estimate.

5 Experiment 4 – If You Use It_2, You'll Lose It_1? How the Use of Different Response Modes in Advice Taking Impacts the Presence and Magnitude of Hindsight Bias

5.1 Introduction

The results of the previous experiments suggest that an active engagement with numerical information is necessary to trigger processes of knowledge revision. In this experiment, I test this idea in the context of hindsight bias. Hindsight bias refers to the finding that the presentation of outcome feedback causes people's recollections of their prior estimates to be biased in the direction of the outcome.

As already mention in Chapter 1, it is commonly assumed that hindsight bias reflects the process of knowledge updating. Specifically, the theoretically most advanced account of hindsight bias in quantitative estimation is the SARA model developed by Pohl and colleagues (Pohl et al., 2003). SARA provides accounts of both anchoring and hindsight bias and asserts that *selective activation* of anchor-/outcome consistent knowledge underlies both biases. Furthermore, the model incorporates an additional mechanism for hindsight bias, *biased sampling*.

According to SARA, the numerical estimation process operates as follows. The knowledge base is assumed to consist of an associative network of units of item-specific knowledge (image set). Furthermore, it is assumed that the content of an image can be transformed into a numerical value. The generation of an estimate consists of the sampling of images. Once a set of images is sampled, the numeric content of the images is averaged. This average corresponds to the estimate.

If the solution to a question is learned (in the hindsight bias task), its value is encoded into the knowledge base. During this process, images that are numerically close to the solution are selectively activated (i.e., the retrieval probability of these images increases). At the recall test, two things can happen. First, if the activation level of the original estimate is high enough, it is retrieved and given as the response. This case would represent a correct recollection. Second, if the original estimate cannot be found, an attempt is made to reconstruct it. Reconstruction is, in general, assumed to be the more frequent process used in the recall task. The reconstruction process involves the same mechanism as the generation of an estimate, namely the sampling (and averaging) of images. Hindsight bias occurs because (a) selectively activated (and outcome-congruent) images are more likely to be sampled, and (b) the provided feedback value can serve as a retrieval cue and bias memory search such that images numerically close to the outcome are more likely to be sampled. Given this biased sample of images, when averaged, the resulting (reconstructed) estimate will be closer to the outcome than the original estimate.

In sum, SARA asserts that hindsight bias results from knowledge updating (Fischhoff, 1975; Hoffrage et al., 2000), biased reconstruction (Stahlberg & Maass, 1998), or the interplay of the two. Critically, these mechanisms are assumed to be driven by automatic processes: "SARA focuses on the automatic processes involved in producing hindsight bias. The model thus reflects the basic observation that the anchoring effect and hindsight bias are extremely robust and can hardly be influenced intentionally" (Pohl et al., 2003, p. 535).

In the present experiment the automatic encoding hypothesis incorporated in SARA—namely that the numerical feedback is automatically encoded into the knowledge base and selectively activates feedback-consistent knowledge—was tested using a hybrid advice-taking/hindsight bias task. Participants first provided estimates for a heterogeneous set of estimation questions. Then they were given the opportunity to revise their judgments in response to information about the estimates of another student. Finally participants were asked to recall their initial estimates.

The results of Experiments 1-3 suggest that advice utilization will result in knowledge revision. That is, if people adopt or assimilate, they should be less likely to correctly recall their prior beliefs. The exact mechanism causing this decrease (e.g., trace substitution, retroactive interference, or biased reconstruction) will not be explored here. In contrast, if people reject the advice, then their knowledge should not be updated. That means, rejectors should be more likely to retain (and access) the trace for the initial estimate. Furthermore, if the initial estimate cannot be recalled, it should be possible to reconstruct it in an unbiased manner.

In addition, the experiment included a source credibility manipulation ("good" advisor vs. unspecified) as well as a manipulation of informational context (advice only vs. advice and prior estimate). These were manipulated because the current experiment used a different set of stimuli and I wanted to determine if the previously observed effects would replicate. Second, these manipulations might also be relevant for the hindsight bias component of the experiment (a) because the re-presentation of the prior estimate during advice taking should increase the probability that the estimate will be correctly recalled subsequently, and (b) because prior studies have indicated that the labeling of the feedback as the "solution" or as "another participant's estimate" did not change the size of the aggregate hindsight bias[18] (Pohl, 1998). Yet, these studies employed the standard hindsight bias task, and not the current hybrid paradigm.

5.2 Method

5.2.1 Participants

One hundred seventy-six introductory psychology students participated in

[18]Interestingly, the advisor label combined with extremely implausible feedback values eliminated the hindsight bias (Pohl, 1998; Hardt & Pohl, 2003). Pohl et al. (2003) acknowledge that their model cannot account for this finding, however, it remains unclear what mechanism could be added to the existing model to counteract the assumed automatic processes.

the study in exchange for partial course credit. All participants were born in Canada and had English as their native language.

5.2.2 Design and Materials

The stimulus set consisted of 24 real-world estimation questions. Numerical advice values were computed on the basis of previously obtained norming data (*N*s ranged from 63 to 81 estimates per question). For each question, the advice value corresponded either to the 30th or 70th percentile of the distribution of normative estimates; half of the questions were randomly assigned the low value (30th percentile) and the other half the high value (70th percentile). Advice values greater than 60 were rounded to reflect people's tendency to use round numbers in estimation tasks (Albers, 2001). The questions and numerical advice are listed in Appendix C.

The stated source credibility of the advisor ('good' vs. unspecified) as well as the information that was presented during the second estimation task were manipulated between subjects. In the 'good' advisor conditions, the alleged advisor was introduced as follows:

> For the purposes of this task, you have been paired with another University of Alberta student. This student was selected from all of those who had participated in a previous real-world estimation study conducted in this lab because this student's performance was very good (that is, level of accuracy was within the top 10% of all participants).

In the unspecified advisor conditions, the last clause specifying the accuracy of the advisor was omitted from the instructions. The informational context of the second estimation task was manipulated such that participants were either presented with their initial estimate and the numerical advice, or with the numerical advice only. Furthermore, the instructions for the advice taking phase of the experiment encouraged participants to use the advice as they saw fit:

> Again, try to come up with the best estimate that you can. If you think it is necessary to revise your initial estimate, feel free to do so, but if you consider your initial estimate still to be the best response, then simply re-enter your initial estimate.

5.2.3 Procedure

Participants were tested individually and all responses were collected by a computer. The experimental session consisted of 3 phases. In the first phase, participants were presented with the 24 estimation questions, one at a time, and were asked to estimate the true target value to the best of their ability. Participants entered their numerical response on the keyboard and confirmed it by pressing the *enter* key. Then, a 5-point rating scale was displayed below the question and the participant's numerical estimate, and the participant had to rate the confidence in the accuracy of their estimate. The rating scale ranged from 1 (*not confident at all*) to 5 (*very confident*). After entering and confirming the rating, the participant proceeded to the next question.

During the second phase, the same 24 questions were presented for a second time. This time, however, each question was presented along with the advisor's estimate and, if applicable, the participant's initial estimate. Again, participants worked their way through the question set providing point estimates and confidence ratings.

The third phase consisted of a surprise recall task. Each question was presented again and participants were asked to recall their initial estimate for that question as accurately as possible. Participants entered the recalled value and confirmed their response by pressing the *enter* key. This caused the program to proceed to the next trial. The presentation order of questions was randomized separately for each participant and each task.

5.3. Results and Discussion

This section is divided into two parts. I will present a response mode analysis of the advice taking data first, and then report the analyses of the recall data.

5.3.1 Advice Taking: Response Mode Analysis and Accuracy Gains

Any responses where the initial estimate was equal to the subsequently presented advice value were removed from the data (200 observations, or 4.7%) because these responses could not be unequivocally assigned to one response mode. Figure 5.1 depicts the WS distributions for each condition. The median WS statistic ranged from .48 when the advisor was presented as highly credible and only the advice was presented, to .75 when the advisor's credibility was left unspecified and both the advice and the participant's prior estimate were present during the target judgment. Thus, the latter condition replicates the standard finding in the advice-taking literature, namely people's tendency to discount the advice of others (Yaniv & Kleinberger, 2000). However, as the multi-modal shape

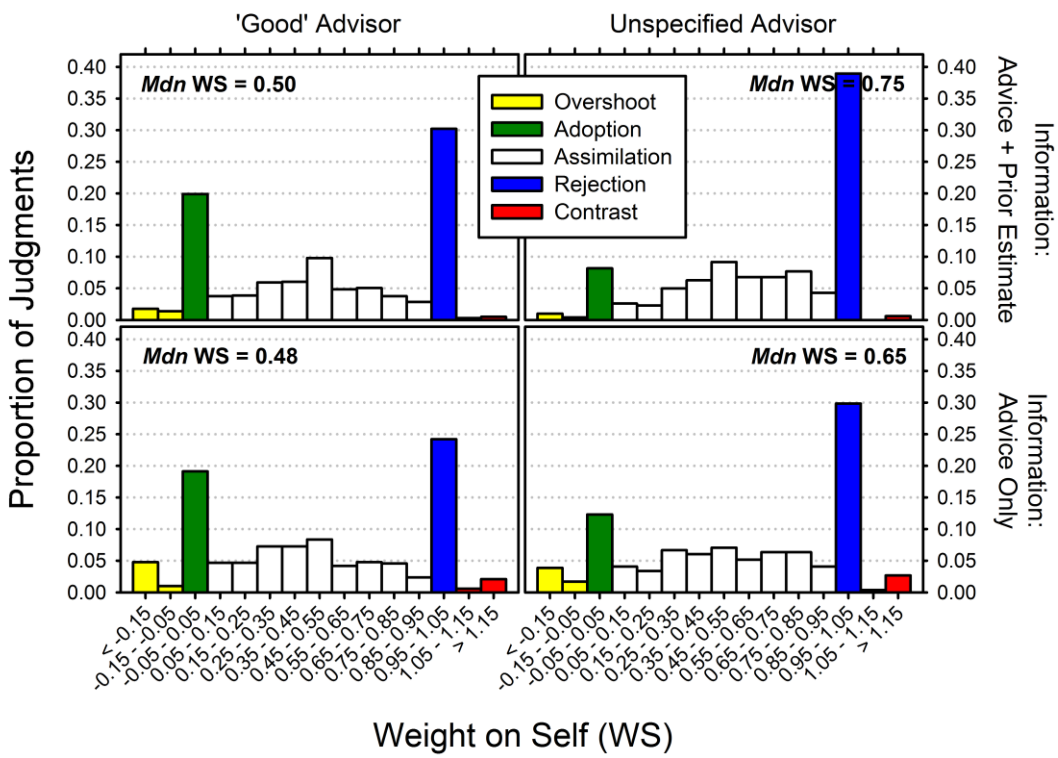

Figure 5.1 Frequency distribution of the binned weight on self (WS) statistic in Experiment 4. The respective response modes (overshoot, adoption, assimilation, rejection, and contrast) are color-coded. Mdn = Median.

of the WS distributions indicates, participants responded to the numerical advice in different ways. Therefore the aggregate WS statistic is of limited use, because it masks the fact that people use different response modes.

In order to determine how the experimental manipulations affected the relative frequency with which the different response modes were used, a multinomial mixed model was fitted to the data. The response modes were treated as a nominal response variable. In order to reduce the number of comparisons and

Table 5.1 Proportions of adoption, rejection, and assimilation responses (in percent) as a function of stated source credibility and information displayed during the advice taking phase (Experiment 4)

Condition	Adoption	Rejection	Assimilation
'Good' Advisor			
Advice + Est_1	20.7	31.4	47.8
Advice only	20.9	26.4	52.7
Unspecified Advisor			
Advice + Est_1	8.3	38.8	51.9
Advice only	13.5	32.6	53.9

Note. The proportions are based on a reduced data set that excluded overshoot and contrast responses; Est_1 = initial estimate.

also due to the low frequency of overshoot and contrast responses, the analysis was restricted to adoption, rejection, and assimilation responses (i.e., overshoot and contrast responses were removed, 5.8% of the data). The absolute proportions of adoption, assimilation, and rejection responses are reported in Table 5.1. The mixed model included subjects and items as random effects and source credibility (good vs. unspecified) and the presence of prior estimate (present vs. absent) as fixed effects. Initial analyses indicated that the source credibility × presence of prior estimate interaction was not significant, $F(2, 3439) = 2.07$, $p = .13$, and the interaction term was therefore removed from the model. The odds ratio estimates are reported in Table 5.2.

As this table indicates, the source credibility manipulation primarily affected the relative rate of adoption and the presentation of the prior estimate impacted the relative rate of rejection. First, the odds that participants rejected rather than adopted the advice (conditional on not assimilating) were smaller when the advice was coming from a source that was presented as highly credible

Table 5.2 Estimated odds ratios (OR) and 95% confidence intervals for the fixed-effects predictors from the multinomial mixed model fitted to the advice-taking data in Experiment 4

Variable	Rejection vs. Adoption		Assimilation vs. Adoption		Rejection vs. Assimilation	
	OR	95% CI	OR	95% CI	OR	95% CI
Source ('Good' vs. Unspecified)	0.43***	[0.31,0.58]	0.51***	[0.38,0.69]	0.83	[0.64,1.07]
Prior estimate (Present vs. Absent)	1.51**	[1.10,2.06]	1.15	[0.84,1.54]	1.30*	[1.01,1.67]

****p* < .001, ***p* < .01, **p* < .05

compared to a source with unknown credibility (OR = 0.43, $t = -5.34$, $p < .001$). Furthermore, the odds of assimilation rather than adoption (conditional on not rejecting) also decreased as source credibility increased (OR = 0.51, $t = -4.47$, $p < .001$). In other words, people tended to adopt advice (rather than assimilate to it or reject it) more frequently when the credibility of the advice was high.

Second, the odds of rejection (rather than adoption) increased by a factor of 1.5 when the prior estimate was presented along with the advice compared to when it was not presented (OR = 1.51, $t = 2.58$, $p = .01$). Similarly, the odds of rejection (rather than assimilation), conditional on not adopting, increased by a factor of 1.3 when the prior estimate was present rather than absent (OR = 1.30, t = 2.02, p = .04). Thus, people tended to maintain their initial beliefs (rather than adopting or assimilating towards the advice) more frequently when their prior estimate was presented along with the advice.

Thus, these results are consistent with findings from Experiments 1-3, albeit the effect of informational context was much smaller with the present paradigm and materials.

Table 5.3 Order of magnitude error (OME) of pre- and post-advice estimates in Experiment 4

Source credibility	Advice + prior estimate				Advice only			
	Initial estimate		Final estimate		Initial estimate		Final estimate	
	M	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
'Good'	0.44	[0.40, 0.49]	0.27	[0.25, 0.29]	0.44	[0.40, 0.47]	0.27	[0.25, 0.28]
Unspecified	0.44	[0.40, 0.47]	0.31	[0.28, 0.33]	0.45	[0.41, 0.49]	0.30	[0.28, 0.32]

Note. CI = 95% confidence interval of the mean.

Accuracy Gains. The metric accuracy (as measured by OME) of post-advice estimates increased in all conditions (see Table 5.3). The results of a mixed ANOVA with OME as the dependent variable revealed a significant effect of estimate (pre-advice vs. post-advice), $F(1, 172) = 320.37$, $p < .001$, partial $\eta^2 = .65$. The estimate × source credibility interaction was marginally significant, $F(1, 172) = 3.38$, $p = .07$, partial $\eta^2 = .02$, reflecting the somewhat greater accuracy gain when the advisor was presented as credible compared to when the credibility of the advisor was left unspecified. All other effects were not significant (*F*s < 1.62).

Given that the advice corresponded to the 30th and 70th percentiles of a large sample of normative estimates, these numerical values therefore tapped into the 'wisdom of crowds' (Larrick, Mannes, & Soll, 2012). And although participants in this study, at least on an aggregate level, tended to egocentrically discount the advice, its influence was still sufficient to improve the overall accuracy of the judgments.

5.3.2 Hindsight Bias: Accuracy and Bias of Recalled Prior Estimates

As for the response mode analysis, the analysis of the recall data was performed on the reduced data set that excluded overshoot and contrast responses. I will first report an analysis on participants' overall recall accuracy. In a second analysis, recall performance will then be analyzed as a function of the response mode used during the advice taking task. Finally, I will present analyses of the bias of incorrectly recalled estimates.

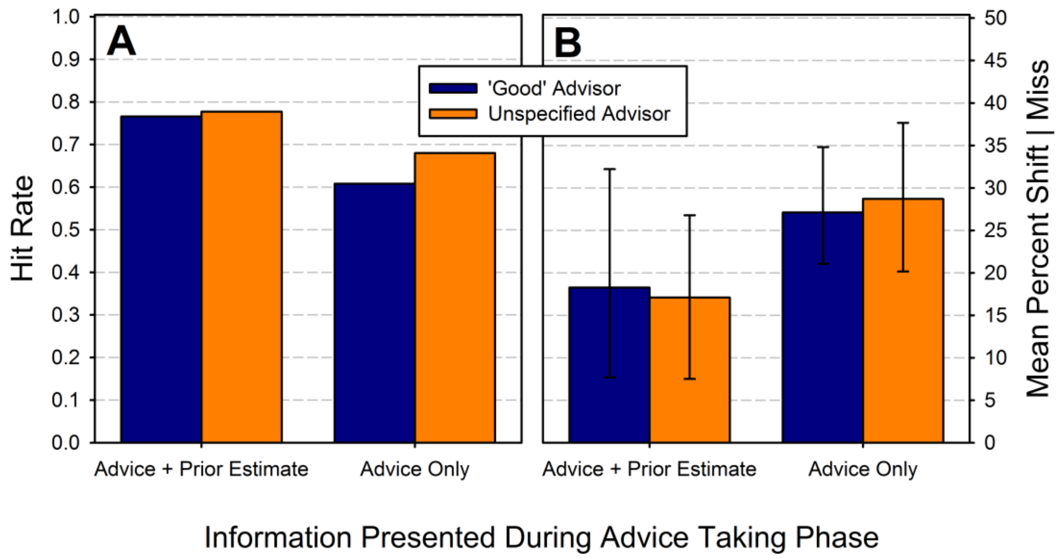

Figure 5.2 Accuracy and bias measures for responses obtained in the recall task in Experiment 4. Panel A shows the aggregate hit rate (i.e., correct recall of initial estimate) for each condition. Panel B depicts the aggregate bias of the incorrectly recalled initial estimates. Percent shift is an index of the extent to which failed recall attempts shifted towards the advisor's estimate. For each condition, the mean of the median by-subject percent shift is shown. Error bars represent 95% bootstrap confidence intervals.

Hit Rate. The proportions of correctly recalled prior estimates are shown in Panel A of Figure 5.2. A logistic mixed model analysis indicated that recall performance was better when the prior estimate had been presented along with the advice, compared to when it had not been presented (OR = 2.10, 95% CI [1.62, 2.72], $t = 3.18$, $p = .002$). This should not be surprising as the additional presentation will most likely lead to a strengthening of the memory trace for this information. Furthermore, the hit rate was slightly lower when the advice came from the ‘good’ advisor rather than the unspecified advisor (OR = 0.82, 95% CI [0.63, 1.06], $t = -1.89$, $p = .06$). The interaction was not significant (log odds = 0.28, $t = 1.07$, $p = .28$).

To determine whether the hit rate was influenced by how participants interacted with the advice, the data were decomposed by response mode (see Figure 5.3). This decomposition revealed that (a) the hit rate for rejected advice was generally very high and was not influenced by whether or not the initial estimate had been present during advice taking, or the credibility of the advisor. (b) In contrast, when participants used the advice in some form (i.e, adoption or assimilation), the hit rate was generally lower, and in both cases the additional presentation of the prior estimate during the advice taking phase did have the beneficial effect seen in the aggregate data. These observations were supported by a logistic mixed model that was fitted to the data (see Table 5.4). Thus, these results suggest that if numerical advice is utilized, it is less likely that prior knowledge states can be accessed subsequently. In contrast, this is not the case if the advice is rejected. The former result suggests a dynamic updating process where outdated beliefs become inaccessible or are eliminated. The latter result reflects that people have some level of control over the management of the knowledge base, and it is inconsistent with an automatic encoding view (Pohl et al., 2003).

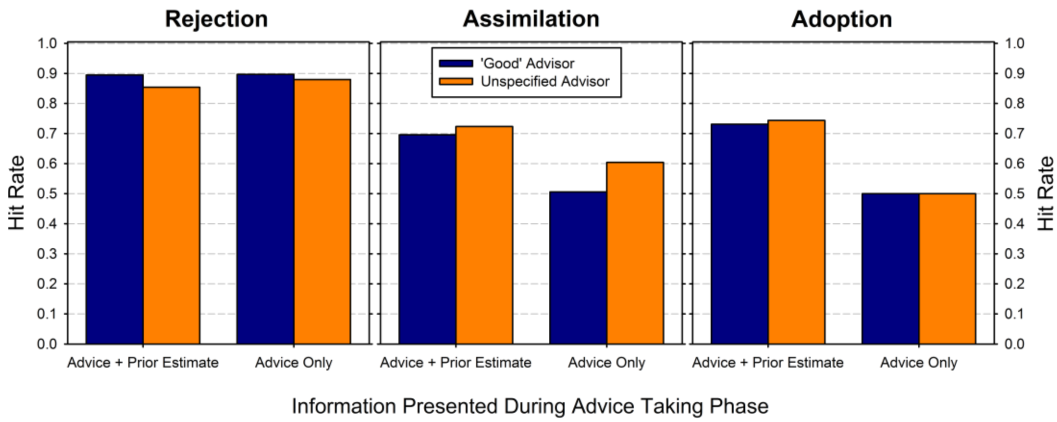

Figure 5.3 Hit rate in the recall task as a function of response mode and advice taking condition.

Table 5.4 Fixed-effects coefficients from the logistic mixed effects model fitted to the recall data in Experiment 4

Variable	Coefficient	*SE*	*t*	*p*
Intercept	2.17	0.20	10.96	<.001
Source-Good	-0.15	0.13	-1.17	.24
Prior Estimate-Present	-0.10	0.21	-0.46	.65
Response Mode-Adoption	-2.02	0.20	-10.34	<.001
Response Mode-Assimilation	-1.77	0.16	-10.88	<.001
Prior Estimate-Present: Response Mode-Adoption	1.31	0.28	4.70	<.001
Prior Estimate-Present: Response Mode-Assimilation	0.88	0.22	4.03	<.001

Note. The model included Subjects and Items as random effects, Source, Prior Estimate, Response Mode, and the Prior Estimate × Response Mode Interaction as fixed effects; the respective reference levels were: Source = Unspecified, Prior Estimate = Absent, Response Mode = Rejection.

Misses. To determine the extent to which incorrectly recalled prior estimates were biased towards the advice values, the percent shift index (PS) was computed for each miss. Percent shift is defined as follows (Hell et al., 1988):

PS = 100 * (Initial Estimate-Recalled Value)/(Initial Estimate-Information)

This measure states the extent (in percent) to which the recalled value shifted towards the information value (in this case the advice). Because the PS index can easily take on extreme values, only responses were analyzed that had a PS value between -250 and 250 (Erdfelder, & Buchner, 1998; Pohl, 2007). This reduced the number of observations (i.e., misses) from originally 1100 to 1043 (5.2% of the data). For each participant, the median PS was computed. Figure 5.2 (Panel B) shows the mean of these medians for each condition. As this graph indicates, the average PS was, in each case, greater than 0 (as indicated by the 95% bootstrap confidence intervals). That is, these responses exhibit a robust hindsight bias.

But does response mode affect the extent to which incorrectly recalled estimates shifted towards the advice? Panel A of Figure 5.4 shows the answer. For this analysis, a participant's median PS was computed separately for each response mode. Because the number of observations per subject and response mode was rather small, the median of the median PS is plotted in Figure 5.4 Panel A. Again, the dissociation between rejection on the one hand, and adoption and assimilation on the other was obtained. When the advice was rejected and the prior estimate was incorrectly recalled, the aggregate bias of these responses (as measured by PS) was small and did not, on an aggregate level, differ from 0

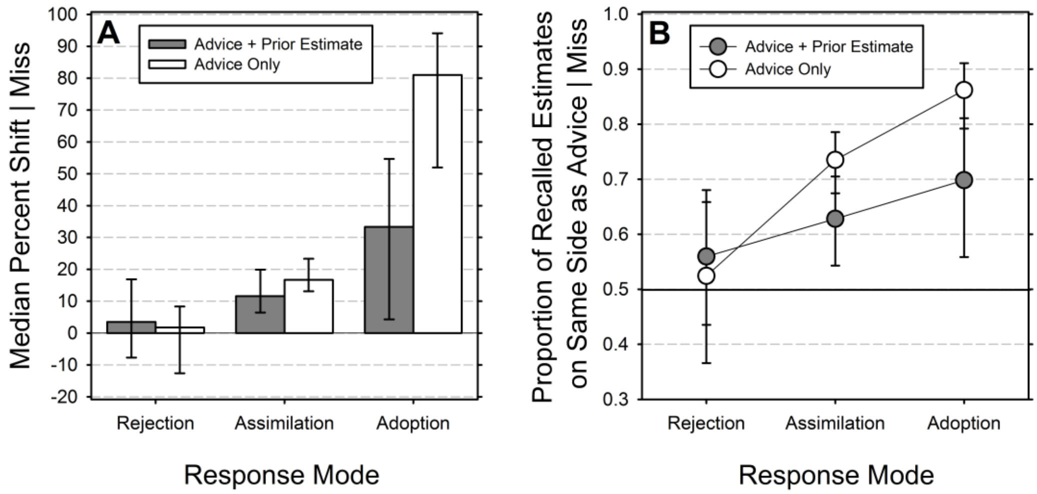

Figure 5.4 Bias measures for incorrectly recalled initial estimates as a function of response mode. Panel A shows the median of the median by-subject percent shift of the recalled estimate towards the advice for each response mode. A value of 0 indicates no bias; values greater than 0 and less than 100 indicate a shift towards the advice; a value of 100 indicate that the recalled estimate equals the advice. Panel B depicts the mean by-subject proportion of recalled estimates located on the same side as the advice relative to the actual initial estimate. Here the baseline is 0.5, which would reflect the absence of a bias. Error bars represent 95% bootstrap confidence intervals.

(representing no bias). In contrast, when participants used the advice in their final judgment (i.e., adoption or assimilation), and then incorrectly recalled their prior estimates, these recalled estimates were biased towards the advice. And, this bias was, at least numerically, somewhat greater when the advice was adopted rather than assimilated to.

In order to obtain converging evidence for this dissociation, a second measure of bias was used. Panel B in Figure 5.4 shows the proportion of responses where both the advice value and the incorreclty recalled value fall on the same side of the actual initial estimate. If one assumes a reconstruction process that operates without any systematic biases, one might expect that half of the incorrectly reconstructed values fall above the target value, and the other half fall below it. Thus, a proportion of 0.5 would respresent the baseline (no bias). Consistent with the relative shift analysis, this bias measure also indicates that the rejection response mode was associated with minimal bias, whereas the adoption and assimilation response modes were associated with a bias greater than 0.5, a bias in the direction of the previously used numerical advice.

5.4 Conclusions

The present findings, in general, are consistent with a controlled view of information uptake. The absence of hindsight bias when the advice was rejected is incompatible with the position that numerical information is automatically encoded into the knowledge base (as, for example, asserted by the SARA model; Pohl et al., 2003). If hindsight bias is indeed a by-product of knowledge updating, then present results suggest that knowledge updating is dependent on the active use of the new information.

However, I need to point out several limitations of the present study. First, the delay between the advice-taking and recall tasks was much shorter in the current study compared to the typical hindsight bias experiment. This was, for example, reflected in high hit rates. Therefore, future studies have to determine if this pattern holds up if the delay is increased. Preliminary data from a follow-up study indicate that the effects replicate if a 30-minute filled delay is used.

A second limitation of this study was that it is unclear what processes (e.g., % reconstruction) people used to recollect their prior estimate. This information, which might be obtained with a strategy menu, could illuminate the mechanisms that drive the hindsight bias in the present paradigm.

6 Conclusions

The central claim of the present thesis is that people have a great deal of cognitive control over how they manage numerical information in their physical and social environments. This claim was supported by several lines of evidence. First, all four experiments demonstrated that people react differently to the numerical information they were exposed to. The use of different response modes was systematic and predictable. For example, information from a highly credible source was generally adopted, whereas information from a source of unknown credibility was occasionally rejected. Second, studies of two different phenomena associated with knowledge revision (i.e., seeding effects and hindsight bias), produced converging evidence that the integration of new information and the revision of existing beliefs requires the active engagement with the new information. That is, people have to actively make use of the numerical information for it to alter the underlying knowledge base.

In the case of seeding effects (or specifically the transfer of advice), Experiments 2-3 demonstrated that people who frequently rejected advice also showed no indication of (metric) knowledge revision (reflected by an absence of transfer). In contrast, those people who did use the advice (without necessarily adopting it), consistently exhibited transfer (in particular if the advice suggested that their initial beliefs were inaccurate).

Similarly, in the case of hindsight bias, the indicator of knowledge updating (i.e., hindsight bias) was only observed when the advice was used, but hindsight bias was eliminated when the advice was rejected. This suggests that if people reject the numerical information, this information does not get automatically integrated into the knowledge base, and thus allows people to retrieve their prior estimates accurately. In contrast, if people use the numerical information in a judgment, this information seems to alter the underlying knowledge base, making it less likely that a prior estimate is reproduced accurately. Although the specific mechanism that underlies the hindsight bias effect in the hybrid advice-taking/hindsight bias paradigm is not well understood yet, the basic finding suggests that models of hindsight bias that assert that the numerical information provided in the task triggers automatic processes that modify the underlying knowledge base (e.g., Pohl et al., 2003) would have to be revised to accommodate the present finding.

In general, the claim that people have the ability to reject numerical information that they deem irrelevant seems obvious. However, a large body of research calls this common-sense view into question. In the literature on judgment under uncertainty, anchoring and hindsight bias are two examples where phenomena have been characterized as virtually inevitable, supposedly, due to automatic mechanisms attributed to the cognitive system. The findings of the present experiments therefore directly challenge this view that people's judgments are easily contaminated by all sorts of (irrelevant) information, without their awareness or intent.

Of course, this controlled information processing view begs the question of why, for example, anchoring effects occur. In other words, if people have control, why would their judgments be influenced by randomly generated numbers? To answer this questions, I first need to point out that, unlike the *standard* anchoring effect which is a robust phenomenon, other anchoring demonstrations (e.g., using subliminal anchors, Mussweiler & Englich, 2005; Reitsma-van Rooijen & Daamen, 2006; or having people copy 5 pages of numbers before they answer an estimation question, Wilson et al., 1996) are more problematic because they tend to be hard to replicate, and typically have a small effect size to begin with (Newell & Shanks, 2014a).

So what underlies the standard anchoring effect? My collaborators and I have argued that anchoring can be understood within this overall framework of controlled information processing (Schweickart et al., 2014). Again, the central idea is that people respond differently to the numerical information. For some people, the anchor is likely to be implausible and they reject it and generate an independent estimate. For others, the anchor value might lie slightly outside the plausible range and they might make an adjustment to a more plausible value (this corresponds to the original anchor-and-adjust account by Tversky and Kahneman, 1974). Finally, for several people the anchor will fall inside the plausible range of values and people will most likely have to guess whether the true value is above or below the anchor. Once they have made this judgment, they provide an estimate that is consistent with this judgment (which typically consists of an adjustment to the next round number). Given these different processes, it is therefore critical to decompose the aggregate anchoring effect to separate the contributions of the different processes to the overall effect.

In one study, we modified the standard anchoring task (comparative judgment followed by estimate) so that the judgment preceding the estimation question consisted of an evaluation of the given anchor value ("Do you consider this number to be a good or bad estimate of the answer to the question?"). When the estimation responses were analyzed as a function of the evaluative response (good vs. bad), a large anchoring effect was observed when participants indicated that the anchor value was a good estimate of the true value. In contrast, a virtually non-existent anchoring effect was observed when participants indicated that the anchor value was a bad estimate of the true value. The elimination of the anchoring effect again supports the view of controlled information processing. From this perspective anchoring effects emerge because for a subset of people, the anchor value happens to fall into a range of metric indifference. Therefore, these people have to guess whether the true value is above or below the anchor, and once this judgment has been made, they (minimally) adjust in the indicated direction. Thus, as long as the anchor value falls into the range of metric indifference for at least a subset of people, an anchoring effect will emerge.

The reported findings on seeding/advice-taking, hindsight bias, and anchoring all suggest that (numerical) information uptake is generally driven by controlled processes. This insight might also help to understand the paradox of why people in general have limited knowledge of quantitative target domains, even though they are constantly exposed to numerical information in their everyday-lives. If indeed an active engagement with the information is a necessary condition for knowledge revision, then simple exposure to the information will not be enough to trigger these effects. This conclusion is also consistent with applied research in cognitive psychology, indicating that simple exposure and repetition generally do not lead to the encoding of complex information (e.g., Bekerian & Baddeley, 1980).

Given that the active engagement with the numerical information seems to be critical to trigger the (potentially) beneficial effects of knowledge revision, this finding also has direct educational implication. For example, the seeding task might serve as an effective tool to improve people's knowledge of various economically- and socially-relevant quantitative content domains (Brown & Siegler, 1993; Wohldmann, 2015).

In conclusion, the integrative approach promoted in this thesis to understand numerical information uptake hopefully spawns new ways of looking at well-known phenomena in the literature on judgment under uncertainty. Of course, at this stage, the framework I outlined requires further empirical tests to determine the extent to which the reported findings generalize and replicate. Furthermore, on a theoretical level, future research should be directed at moving from a descriptive level (e.g., the response mode terminology) to an explanatory one (e.g., linking processes to response modes). Finally, the current framework and findings might serve to somewhat tone down the often overly negative portrayals of human judgment as susceptible and bias-prone.

Bibliography

Albers, W. (2001). Prominence theory as a tool to model boundedly rational decisions. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 297–317). Cambridge, MA: MIT Press.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). “Coherent arbitrariness”: Stable demand curves without stable preferences. *The Quarterly Journal of Economics, 118*, 73-105.

Arkes, H. R. (1991). Costs and benefits of judgment error: Implications for debiasing. *Psychological Bulletin, 110*, 486-498. doi:10.1037/0033-2909.110.3.486

Ash, I. K. (2009). Surprise, memory, and retrospective judgments making: Testing cognitive reconstruction theories of the hindsight bias effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 916-933. doi:10.1037/a0015504

Bekerian, D. A., & Baddeley, A. D. (1980). Saturation advertising and the repetition effect. *Journal of Verbal Learning and Verbal Behavior, 19*, 17-25. doi:10.1016/S0022-5371(80)90476-4

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48-74. doi:10.1037//0022-3514.37.1.48

Birnbaum, M. H., Wong, R., & Wong, L. K. (1976). Combining information from sources that vary in credibility. *Memory & Cognition, 4*, 330-336. doi:10.3758/BF03213185

Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition, 25*, 132-146. doi: 10.1521/soco.2007.25.1.132

Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition, 106*, 1408-1440. doi:10.1016/j.cognition.2007.07.007

Blankenship, K. L, Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology, 44*, 1465-1476. doi:10.1016/j.jesp.2008.07.005

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*, 127-151. doi:10.1016/j.obhdp.2006.07.001

Brewer, N. T., & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making, 15*, 65-77. doi:10.1002/bdm.403

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 321-360). New York, NY: Academic Press.

Brown, N. R., Cui, X., & Gordon, R. D. (2002). Estimating national populations: Cross-cultural differences and availability effects. *Applied Cognitive Psychology, 16*, 811-827. doi:10.1002/acp.830

Brown, N. R., & Siegler, R. S. (1992). The role of availability in the estimation of national populations. *Memory & Cognition, 20,* 406-412. doi:10.3758/BF03210924

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100*, 511-534. doi:10.1037//0033-295X.100.3.511

Brown, N. R., & Siegler, R. S. (1996). Long-term benefits of seeding the knowledge base. *Psychonomic Bulletin & Review, 3*, 385-388. doi:10.3758/BF03210766

Brown. N. R., & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition, 29*, 405-412. doi:10.3758/BF03196391

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7*, 223-242. doi:10.1002/bdm.3960070402

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes, 79*, 115-153. doi:10.1006/obhd.1999.2841

Chaxel, A.S. (2014). The impact of procedural priming on selective accessibility on self-generated and experimenter-provided anchors. *Journal of*

Experimental Social Psychology, 50, 45-51. doi:10.1016/j.jesp.2013.09.005

Central Intelligence Agency. (2011). Country comparison: Population. In *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2119rank.html

Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48*, 147-168. doi:10.1016/0749-5978(91)90010-Q

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making, 21*, 241-251. doi: 10.1002/bdm.586

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and business. *Psychological Sciences in the Public Interest, 5*, 69-106. doi:10.1111/j.1529-1006.2004.00018.x

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts‘ judicial decision making. *Personality and Social Psychology Bulletin, 32*, 188-200. doi:10.1177/0146167205282152

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12*, 391-396. doi:10.1111/1467-9280.00372

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30*, 447-460. doi:10.1177/0146167203261889

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. doi:10.1002/bdm.495

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*, 311-318. doi:10.1111/j.1467-9280.2006.01704.x

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 387-414. doi:10.1037/0278-7393.24.2.387

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278. doi: 10.1146/annurev.psych.59.103006.093629

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 288-299. doi:10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 349-358. doi:10.1037/0096-1523.3.2.349

Fischhoff, B., & Beyth, R. (1975). “I know it would happen.” Remembered probabilities of once-future things. *Organizational Behavior and Human Decision Processes, 13*, 1-16. doi:10.1016/0030-5073(75)90002-1

Friedman, A., & Brown, N. R. (2000a). Reasoning about geography. *Journal of Experimental Psychology: General, 129*, 193-219. doi:10.1037//0096-3445.129.2.193

Friedman, A., & Brown, N. R. (2000b). Updating geographical knowledge: Principles of coherence and inertia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 900-914. doi:10.1037//0278-7393.26.4.900

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*, 35-42. doi:10.1016/j.socec.2010.10.008

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes, 107*, 234-245. doi:10.1016/j.obhdp.2008.03.001

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making, 20*, 21-35. doi:10.1002/bdm.539

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *Journal of Applied Psychology, 93*, 1165-1173. doi:10.1037/0021-9010.93.5.1165

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes, 108*, 287-302. doi:10.1016/j.obhdp.2008.08.002

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York, NY: Academic Press.

Hardin, J. W., & Hilbe, J. M. (2013). *Generalized estimating equations*. Boca Raton, FL: CRC Press.

Hardt, O., & Pohl, R. (2003). Hindsight bias as a function of anchor distance and anchor plausibility. *Memory, 11*, 379-394. doi:10.1080/09658210244000504

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making, 17*, 333-348. doi:10.1002/bdm.474

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*, 117-133. doi:10.1006/obhd.1997.2697

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes, 81*, 252-273. doi:10.1006/obhd.1999.2874

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. doi:10.1037/0033-295X.93.3.258

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*, 311-327. doi: 10.1037/0033-2909.107.3.311

Hell, W., Gigerenzer, G., Gaugell, S., Mall, M., & Müller (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition, 16*, 533-538. doi: 10.3758/BF03197054

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231-237. doi:10.1111/j.1467-9280.2009.02271.x

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 566-581. doi:10.1037//0278-7393.26.3.566

Hogarth, R. M., & 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

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161-1166. doi:10.1177/01461672952111004

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

Koehler, D. J., & Beauregard, T. A. (2006). Illusion of confirmation from exposure to another's hypothesis. *Journal of Behavioral Decision Making, 19*, 61-78. doi:10.1002/bdm.513

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221-232. doi: 10.1037//0022-3514.77.2.221

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227-242). New York, NY: Psychology Press.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*, 111-127. doi:10.1287/mnsc.1050.0459

LaVoie, N. N., Bourne, L. E., Jr., & Healy, A. F. (2002). Memory seeding: Representations underlying quantitative estimations. *Journal of Experimental Psychology: Learning, memory, and Cognition, 28*, 1137-1153. doi:10.1037//0278-7393.28.6.1137

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making, 8*, 149-168. doi:10.1002/bdm.3960080302

Matthews, W. J. (2011). What might judgment and decision making research be like if we took a Bayesian approach to hypothesis testing? *Judgment and Decision Making, 6*, 843-856.

Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes, 102*, 42-58. doi:10.1016/j.obhdp.2006.09.005

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes, 98*, 133-143.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136-164. doi:10.1006/jesp.1998.1364

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*, 1142-1150. doi:10.1177/01461672002611010

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1399-1413. doi:10.1037/a0020449

Newell, B. R., & Shanks, D. R. (2014a). Prime numbers: Anchoring and its implications for theories of behavior priming. *Social Cognition, 32*, 88-108. doi:10.1521/soco.2014.32.supp.88

Newell, B. R., & Shanks, D. R. (2014b). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*, 1-61. doi:10.1017/S0140525X12003214

Nickerson, R. S. (1981). Motivated retrieval from archival memory. In H. E. Howe, Jr. & J. H. Flowers (Eds.), *Nebraska symposium on motivation: Cognitive processes* (Vol. 28, pp. 73-119). Lincoln, NE: University of Nebraska Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220. doi:10.1037/1089-2680.2.2.175

Oppenheimer, D. M., LeBoeuf, R. A., & Brewer, N. T. (2008). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition, 106*, 13-26. doi:10.1016/j.cognition.2006.12.008

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

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205). New York, NY: Academic Press.

Plous, S. (1989). Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology, 19*, 67-91. doi:10.1111/j.1559-1816.1989.tb01221.x

Pohl, R. F. (1998). The effects of feedback source and plausibility of hindsight bias. *European Journal of Cognitive Psychology, 10*, 191-212. doi:10.1080/713752272

Pohl, R. F. (2004). Hindsight bias. In R. F. Pohl (Ed.), *A handbook on fallacies and biases in thinking, judgement and memory* (pp. 363-378). New York, NY: Psychology Press.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition, 25*, 14-31. doi:10.1521/soco.2007.25.1.14

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory, 11*, 337-356. doi:10.1080/09658210244000487

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes, 67*, 49-58. doi:10.1006/obhd.1996.0064

Ravazzolo, F., & Røisland, Ø. (2011). Why do people place lower weight on advice far from their own initial opinion? *Economics Letters, 112*, 63-66. doi:10.1016/j.econlet.2011.03.032

Reitsma-van Rooijen, M., & Daamen, D. D. L. (2006). Subliminal anchoring: The effects of subliminally presented numbers on probability estimates. *Journal of Experimental Social Psychology, 42*, 380-387.

Reyna, V. F., Nelson, W. L., Han, P. K., & Diekmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135*, 943-973. doi:10.1037/a0017327

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7*, 411-426. doi:10.1177/1745691612454303

Russo, J. E., & Schoemaker, P. J. H. (1989). *Decision traps*. New York: Simon and Schuster.

Schultze, T., Rakotoarisoa, A., & Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making, 10*, 144-171.

Schwarz, N. (1994). Judgment in social context: Biases, shortcomings, and the logic of conversation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 123-162). San Diego, CA: Academic Press.

Schweickart, O., Tam, C., & Brown, N. R. (2014). *When “bad” is good: How evaluative judgments eliminate the standard anchoring effect*. Unpublished manuscript.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250-264. doi:10.1037//0096-3445.116.3.250

Smith, A. R., & Windschitl, P. D. (2015). Resisting anchoring effects: The roles of metric and mapping knowledge. *Memory & Cognition*. Advance online publication. doi:10.3758/s13421-015-0524-4

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes, 62*, 159-174. doi:10.1006/obhd.1995.1040

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting, 27*, 81-102. doi:10.1016/j.ijforecast.2010.05.003

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others’ opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 780-805. doi:10.1037/a0015145

Stahlberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory or biased reconstruction? *European Review of Social Psychology, 8*, 105-132. doi: 10.1080/14792779643000092

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*, 437-446. doi: 10.1037/0022-3514.73.3.437

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.

Switzer, F. S., III, & Sniezek, J. A. (1991). Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Processes, 49*, 208-229. doi:10.1016/0749-5978(91)90049-Y

Toglia, M. P., Ross, D. F., Cici, S. J., & Hembrooke, H. (1992). The suggestibility of children's memory: A social-psychological and cognitive interpretation. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *The development of long-term retention* (pp. 217-241). New York, NY: Springer.

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes, 117*, 53-65. doi:10.1016/j.obhdp.2011.10.001

Touloumis, A. (2015). R package multgee: A generalized estimating equations solver for multinomial responses. *Journal of Statistical Software, 64*, 1-14.

Touloumis, A., Agresti, A., & Kateri, M. (2013). GEE for multinomial responses using a local odds ratios parameterization. *Biometrics, 69*, 633-640. doi:10.1111/biom.12054

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

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, 19*, 645-647. doi:10.1111/j.1467-9280.2008.02136.x

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, 37, 62-69. doi:10.1006/jesp.2000.1431

Wiest, L. R., Higgins, H. J., & Frost, J. H. (2007). Quantitative literacy for social justice. *Equity & Excellence in Education, 40*, 47-55. doi: 10.1080/10665680601079894

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117-142. doi: 10.1037/0033-2909.116.1.117

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. doi:10.1037//0096-3445.125.4.387

Wohldmann, E. L. (2015). Planting a seed: Applications of cognitive principles for improving food choices. *The American Journal of Psychology, 128*, 209-218. doi:10.5406/amerjpsyc.128.2.0209

Wong, K. F. E., & Kwong, J. Y. Y. (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes, 82*, 314-333. doi:10.1006/obhd.2000.2900

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes, 69*, 237-249. doi:10.1006/obhd.1997.2685

Yaniv, I. (2004a). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1-13. doi:10.1016/j.obhdp.2003.08.002

Yaniv, I. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science, 13*, 75-78. doi:10.1111/j.0963-7214.2004.00278.x

Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making, 25*, 427-434. doi:10.1002/bdm.740

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 558-563. doi:10.1037/a0014589

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*, 260-281. doi:10.1006/obhd.2000.2909

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103,* 104-120. doi:10.1016/j.obhdp.2006.05.006

Appendices

Appendix A

Confirmatory and Disconfirmatory Sets of Countries Used as Information Items (Population Values [in Millions] are Taken From the World Factbook [Central Intelligence Agency, 2011])

Confirmatory		Disconfirmatory	
Country	Actual population	Country	Actual population
Russia	138.7	Bangladesh	158.6
Thailand	66.7	Myanmar	54.0
Poland	38.4	Tanzania	42.7
Peru	29.2	Iraq	30.4
Taiwan	23.1	Australia	21.8
Niger	16.5	Burkina Faso	16.8
Angola	13.3	Cuba	11.1
Guinea	10.6	Czech Republic	10.2
Haiti	9.7	Austria	8.2
Laos	6.5	Israel	7.5

Appendix B

Table 3.1S Coefficients of the multinomial GEE model fitted to the data (information items) in Experiment 2

Coefficient	Rejection vs. Adoption				Assimilation vs. Adoption				Rejection vs. Assimilation			
	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*	Estimate	*SE*	*z*	*p*
Intercept	-1.11	0.31	-3.55	<.001	0.90	0.22	4.03	<.001	-2.01	0.23	-8.82	<.001
Source (Avr vs. Unsp)	-1.00	0.32	-3.18	.002	-0.59	0.25	-2.37	.02	-0.42	0.23	-1.85	.06
Prior estimate (Pres vs. Abs)	1.88	0.33	5.71	<.001	0.90	0.25	3.57	<.001	0.98	0.24	4.10	<.001
OME_1	-1.12	0.29	-3.83	<.001	0.04	0.17	0.23	.82	-1.16	0.26	-4.46	<.001

Note. Avr = Average, Unsp = Unspecified, Pres = Present, Abs = Absent, OME_1 = order of magnitude error for initial estimate.

Appendix C

Questions and Numerical Advice Used in Experiment 4

Question	Actual value	Numerical advice	Percentile of normative distribution
How many homicides occurred in Edmonton in 2013?	28	50	70th
How many gold medals did China win at the 2012 Summer Olympic Games?	38	30	70th
How old was Neil Armstrong when he landed on the moon?	39	31	30th
What was the highest (hottest) recorded temperature for a day in Los Angeles, California (in degrees Celsius)?	45	53	70th
How many countries are there in Africa?	54	38	70th
What is the average life expectancy of an Asian elephant in the wild (in years)?	70	60	70th
What is the fastest speed a cheetah can run (in km/hour)?	120	65	30th
How long was the movie 'Forrest Gump' (in minutes)?	142	120	30th
How many bones make up a dog skeleton?	319	100	30th
What is the height of the Eiffel Tower (in metres)?	324	1,000	70th
What is the weight of an adult zebra (in kg)?	350	300	70th
How many calories are in a McDonald's 'Big Mac' sandwich?	540	1,000	70th

What is the maximum seating capacity of a Boeing 747 (a jumbo jet)?	660	240	30th
What is the distance between Edmonton and Toronto (in km)?	2,701	2,600	30th
How much does a 2012 Toyota Prius weigh (in pounds)?	3,042	2,000	70th
How many pages are there in the complete Harry Potter book series (UK version)?	3,407	3,000	30th
How long is the Mississippi River (in km)?	3,734	3,300	70th
What is the total student enrollment at the University of Calgary?	32,160	20,000	30th
How many Canadian soldiers died in World War II?	45,400	150,000	70th
What is the base tuition fee at Harvard Law School for the 2013-2014 academic year (in US dollars)?	52,350	20,000	30th
What is the manufacturer's suggested retail price of a 2013 Ferrari FF Sports Car (in US dollars)?	295,000	160,000	30th
How many babies were born in Canada in 2013?	383,822	75,000	30th
What is the annual base salary of the President of the United States (in US dollars)?	400,000	800,000	70th
What was the population of San Francisco, California in 2013?	825,111	2,000,000	30th

Note. Advice values greater than 60 were rounded.