The Effects of Expertise on the Hindsight Bias

A Dissertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Melissa A.Z. Marks Knoll, B.A., M.A.

Graduate Program in Psychology

* * * * *

The Ohio State University
2009

Dissertation Committee:

Dr. Hal R. Arkes, Advisor

Dr. Thomas E. Nygren

Dr. Michael C. Edwards

ABSTRACT

I present data from three experiments in which I explored the effects of expertise on the hindsight bias. In Experiment 1 participants read an essay about baseball or about owning a dog and then answered a 20-question true/false quiz about the baseball essay to the best of their ability (do-your-best group), as if they had not read the essay (discount group), or to the best of their ability even though they read about owning a dog (dogs group). Participants also completed a quiz about baseball rules (measure of expertise). Results demonstrated that as participants' baseball expertise increased, their inability to act as if they had never read the essay also increased; expertise exacerbated hindsight bias. In Experiment 2, varsity baseball players and baseball non-experts answered a 20-question quiz about baseball current events. Foresight participants answered the questions, while hindsight participants were given the questions and the answers and had to give the probability that they would have known the answers had the answers not been provided. The baseball players displayed no hindsight bias, while non-experts demonstrated the bias. To test of the effects of subjective expertise on hindsight bias, participants in Experiment 3 ranked five topics in order of expertise and gave feeling-of-knowing (FOK) ratings for 100 questions from these topics. Foresight participants then saw each question again and answered the questions, while hindsight participants saw the

questions and answers and gave the probability they would have known the answers had they not been provided. Three important results emerged: average FOK increased with subjective expertise; accuracy increased with subjective expertise, but overall performance was poor; hindsight bias increased with subjective expertise. Taken together the results demonstrate that hindsight bias is exacerbated by expertise up to a certain point, after which experts demonstrate no bias. The effects of expertise on hindsight bias can be attributed to variations in feelings of knowing. An inverted U-shaped relationship between expertise and hindsight bias is proposed to explain why naïve individuals and extreme experts fail to show the bias while those with intermediate knowledge do.

For my husband

ACKNOWLEDGMENTS

I want to thank my advisor, Hal Arkes, for his constant encouragement and support through my graduate career. The lessons I have learned from him on both a professional and personal level have been invaluable and will remain with me throughout my life. Hal's patience and respect are rare virtues, and I could not have asked for a more kind and compassionate mentor.

I also wish to thank my family, especially my mom, for supporting me in all of my endeavors. I could not have made it through 23 years of school without your unconditional love and support.

Finally, I want to thank my husband and best friend, Aaron, whose unwavering support and encouragement have allowed me to accomplish all that I set out to do.

The research presented in this dissertation was funded by a grant from the National Science Foundation (SES-0548605).

VITA

Graduate Research Assistant, The Ohio State University.....September 2004 – Present

Graduate Teaching Assistant, The Ohio State University.....September 2004 – Present
Courses: Psychology of Judgment and Decision Making
Introduction to Data Analysis in Psychology

Substitute Lecturer- Judgment and Decision Making.....April 2007 – May 2007

Co-Principal Investigator, Incentive Marketing Association.....April 2007 - July 2007

Market Research Intern, BIGresearch, LLC.....June 2006 – September 2006

PUBLICATIONS

Marks, M. A. Z, & Arkes, H. R. (2008). Patient and surrogate disagreement in end-of-life decisions: Can surrogates accurately predict patients' preferences? *Medical Decision Making*, 28, 524-531.

Marks, M. A. Z, & Arkes, H. R. (published online, 2009). The effects of mental contamination on the hindsight bias: Source confusion determines success in disregarding knowledge. *Journal of Behavioral Decision Making*.

FIELDS OF STUDY

Major Field: Psychology
Areas of Concentration: Quantitative Psychology and Judgment and Decision Making

TABLE OF CONTENTS

	Page
Abstract.....	ii
Dedication.....	iv
Acknowledgments.....	v
Vita.....	vi
List of Tables.....	ix
List of Figures.....	x
INTRODUCTION.....	1
CHAPTER 1: Expertise: Friend or Foe?.....	3
CHAPTER 2: Hindsight Bias and Expertise: Is There a Curse of Knowledge?.....	12
CHAPTER 3: Experiment 1: The Effects of Source Confusion on Hindsight Bias in Experts and Novices.....	31
CHAPTER 4: Experiment 2: Are “Super Experts” Resistant to the Hindsight Bias?.....	42
CHAPTER 5: Experiment 3: Feeling-of-Knowing Explains the Relationship between Expertise and Hindsight Bias.....	54
CHAPTER 6: General Discussion.....	84
LIST OF REFERENCES.....	97
APPENDIX A: Baseball and Dogs Essays and Baseball Quiz.....	104
APPENDIX B: Baseball Expertise Quiz.....	109

APPENDIX C: Baseball Current Events Quiz.....	112
APPENDIX D: Five-Topic General Knowledge Quiz.....	115
APPENDIX E: Tables and Figures.....	123

LIST OF TABLES

Table		Page
1	Regression coefficients from Experiment 1	123

LIST OF FIGURES

Figure		Page
1	Expertise quiz score results for Experiment 1	124
2	Regression graph for Experiment 1	125
3	Expertise quiz score results for Experiment 2	126
4	Relationship between expertise and feeling-of-knowing in Experiment 3	127
5	Relationship between expertise and questions answered correctly in each domain in Experiment 3	128
6	Relationship between expertise and hindsight bias in Experiment 3	129
7	Inverted U-shaped relationship between expertise and hindsight bias	130

INTRODUCTION

My husband and I watch Jeopardy! every night after dinner. As a graduate student in psychology, former Spanish major, and avid cook, I get excited when Alex Trebek introduces categories dealing with language origins, food, and psychology. My husband, an aspiring physical education teacher and English major, perks up when the categories deal with sports or literature. These are the categories about which we think we know the most, and often we shout out the correct answers. However, these are also the categories in which we most often say “Darn, I knew that!” after failing to recall the correct answer and subsequently hearing the answer from Alex or an astute contestant. Is it the case that we really *did* know the answers to the questions in our areas of “expertise,” or did we just *think* or *feel like* we knew the answers? When the category is European literature, I am unfazed when I fail to answer a question and then hear the correct answer, and my husband barely reacts when the same events occur for the category of French cuisine. Why do we not say “Darn, I knew that!” after hearing the correct answer to a question outside of our respective areas of expertise?

In the current research I examine the notion that hindsight bias is affected by objective and subjective expertise. In a series of three experiments, I explore the interplay between expertise and individuals’ tendency to report that they “knew all along”

the answers to questions that they did not in fact know in foresight. Experiment 1 demonstrates that as participants' expertise in baseball increases, so does their inability to disregard information about baseball current events. The methodology implemented in Experiment 1 specifically tests individuals' ability to identify as new or old facts presented in story form; experts in baseball demonstrate greater difficulty in separating newly acquired baseball information from the knowledge possessed prior to participation in the experiment. In Experiment 2 I employed a traditional test of the hindsight bias and recruited Ohio State University varsity baseball players, who may be considered "super experts" in baseball. The baseball players displayed virtually no hindsight bias when they were asked about baseball current events. The use of "super experts" in Experiment 2 adds an interesting wrinkle to the expertise-hindsight bias story; do the most expert individuals know what they do and do not know, effectively shielding them from hindsight bias? Finally, I demonstrate in Experiment 3 that subjective expertise leads to an increased "feeling-of-knowing" in one's area of presumed expertise without leading to a corresponding increase in accuracy. This discrepancy between what people actually know and what they *think* or *feel like* they know results in increased hindsight bias. Finally, I propose that an inverted U-shaped relationship exists between expertise and hindsight bias, which accounts for the fact that naïve individuals and "super experts" do not display hindsight bias, while those possessing intermediate levels of knowledge do fall prey to the bias.

CHAPTER 1

Expertise: Friend or Foe?

Experts are called upon by courts of law, media outlets, and countless other parties in search of definitive answers to difficult questions. Viewers would be hard pressed to accept as fact the opinion of a local news anchor on the costs and benefits of stem-cell research. Jurors would be unlikely to factor into their verdicts statements made by a psychiatric hospital's custodian testifying about the imminent dangerousness of a patient. Experts hold a special place in the minds of lay people, who appreciate the unique knowledge experts possess. But, is expertise always worthy of the acclaim it receives? Shanteau (1992) described a dichotomy in the portrayal of expert competency in the literatures of different branches of psychology, noting that judgment and decision making researchers tend to focus on experts' missteps, while cognitive science researchers paint a rosy picture of experts. Researchers on both sides have explored numerous aspects of expertise, including specific qualities experts exemplify (Ericsson & Charness, 1994), the kinds of strategies experts use to solve problems in their areas of expertise (Larkin, McDermott, Simon, & Simon, 1980), differences in the costs of information search between experts and novices (Camerer & Johnson, 1997), differences in the way experts and novices incorporate new information (Chiesi, Spilich, & Voss

1979), and under what conditions experts are likely or unlikely to perform competently (Shanteau, 1992). As noted above, some researchers have questioned whether expertise is always beneficial, or whether expertise can, in some instances, be detrimental.

Consistent with conventional wisdom, expertise often affords individuals with substantial benefits over novices. For example, Spilich, Vesonder, Chiesi, and Voss (1979) demonstrated that participants high in baseball knowledge (high knowledge; HK) were able to process more readily information presented about events in a baseball game than were participants low in baseball knowledge (low knowledge; LK). Spilich et al. (1979) argued that HK individuals have better knowledge than LK individuals regarding the “goal structure” (e.g., scoring runs, preventing runs from being scored, winning the game), “game actions” (e.g., a pitcher throwing a “strike,” the batter striking out, the left-fielder catching a fly ball), and “sequences of actions” (e.g., the shortstop fielded a “line drive” and threw the ball to the first baseman for a “double play”) that occur in baseball games. Importantly, the authors posited that HK individuals are better than LK individuals at relating game actions and sequences of actions to the goal structure and successfully keeping in working memory “values” of the goal-related game actions (e.g., there are men on first and third base) in a way similar to the play-by-play updates given by game announcers. In order to test these hypothesized differences between HK and LK individuals, Spilich et al. (1979) presented participants with a taped account of a half of an inning of a baseball game and asked participants to a) summarize what happened in the half-inning, b) recount as much of the half-inning as possible, and c) answer 40 questions regarding specific events contained in the half-inning account. The results confirmed the authors’ predictions. HK participants recalled more and answered more

questions correctly about events from the account than did LK participants, and the specific events HK participants recalled were more relevant to the goal structure of the game. Specifically, HK participants recounted more events related to game actions that actually resulted in important changes in the goal structure of the game, while LK participants more often recounted game actions that were unrelated to the goal structure. Spilich et al. also found HK participants' statements to be more embellished than statements written by LK participants; HK participants made more inferences about the people and events described in the half-inning account (e.g., describing a pitcher identified only as "left-handed" and capable of striking out "quite a few batters" in the account as a "big, fastballing lefthander"). In general, Spilich et al. (1979) found evidence that high-knowledge individuals have an advantage over low-knowledge individuals in terms of the incorporation and recall of relevant, goal-related information.

Larkin et al. (1980) also discussed the benefits of expertise, focusing on the ways in which experts represent information in long-term memory, their perceptual knowledge, and their recognition capabilities in their areas of expertise. Larkin et al. (1980) described the typical experiment used to demonstrate expert performance in chess players (e.g., Chase & Simon, 1973; deGroot, 1965; Simon & Chase, 1973). In this paradigm, participants are shown a chess board containing a position with approximately 25 pieces on the board. After seeing the chess board for about 5-10 seconds, participants are asked to recreate the position from memory. While chess masters and grand masters perform this test with extremely high accuracy (approximately 90% accuracy), less-skilled players typically remember the placement of about 5 or 6 pieces. In the second phase of the experiment, pieces are randomly placed on the chess board, and the memory task is

repeated. In this case, masters and grand masters actually perform worse than do the novice players. In the latter case, expert chess players fail to remember the placement of the pieces because they are unable to exploit the skill afforded by their expertise, namely their capacity to recognize and store “chunks” of multiple pieces set up in typical configurations. A chunk is a stimulus composed of multiple units that, due to repeated exposure, have come to be recognized as a single entity. While novices can only hold in short-term memory about 5 or 6 pieces at a time, experts can hold in short-term memory about 4 times as many pieces, since they can remember chunks of pieces in lieu of single chess pieces. When the pieces are randomly placed on the chess board, experts must use the same strategy as novices when memorizing the placement of the pieces, thus eliminating any advantage expertise typically provides. This demonstration illustrates a benefit of expertise: experts are able to recognize typical configurations of information in their area of expertise. Furthermore, experts store or “index” these chunks of information in long-term memory and are able to access them when necessary. In this sense, chess masters are similar to the high knowledge participants in the Spilich et al. (1979) experiment described above. Participants high in baseball knowledge had in memory representations of baseball goal structures and sequences of actions typically observed in baseball games. When they were presented with baseball information, HK participants were able to act similarly to chess masters recreating a chess position from memory: HK participants remembered more from the half-inning account than did LK participants because HK participants could remember sequences of actions in “chunks” typical in baseball games.

As Shanteau (1992) noted in his discussion of expert competence, experts are not infallible. Examples of errors in expert decision making abound in the judgment and decision making literature and often are explained in the “heuristics and biases” (Tversky & Kahneman, 1974) tradition. For example, Poses and Anthony (1991) examined erroneous probability estimates made by physicians and implicated inappropriate use of the availability heuristic (Tversky & Kahneman, 1974) and wishful thinking in physicians’ overdiagnosis of bacteremia in hospitalized patients. Casscells, Schoenberger, and Graboyes (1978) found that most of the 20 house officers, 20 attending physicians, and 20 fourth-year medical students polled were unable to arrive at anything near the correct likelihood (i.e., the correct Bayesian posterior probability) of a patient having a particular disease, even though they were given the data necessary to arrive at the correct answer. Caplan, Posner, and Cheney (1991) demonstrated that when anesthesiologists were given a scenario in which exactly the same standard of care was adopted by a doctor during intubation, but a negative outcome resulted that was either permanent or temporary, the anesthesiologists judged the doctor to have acted more inappropriately when the outcome was permanent than when the negative outcome was temporary. These are only a sampling of the many examples of physicians’ tendency to succumb to cognitive biases (e.g., Chapman & Elstein, 2000; Kern & Doherty, 1982; McNeil, Pauker, Sox, & Tversky, 1982; Redelmeier & Shafir, 1995; Tversky & Kahneman, 1986).

In addition to examples of physicians displaying cognitive biases, judges have also been shown to demonstrate cognitive errors within their own profession. Guthrie, Rachlinski, and Wistrich (2001) tested judges’ susceptibility to anchoring, framing

effects, hindsight bias, egocentric bias, and use of the representativeness heuristic and found that judges are subject to all of these cognitive errors. For example, two groups of judges were given a scenario regarding a personal injury lawsuit and were asked to state how much they would award the plaintiff in compensatory damages. The group of judges who saw an additional, but invalid, argument from the defendant that the case should be dismissed because it did not meet the \$75,000 jurisdictional minimum said they would award the plaintiff \$367,000 less than judges in the group who did not receive the defendant's argument. Even though the defendant's argument was invalid because the plaintiff clearly suffered damages far in excess of \$75,000, and even though judges recognized that the defendant's argument had no merit (only 2% of judges said they would grant the motion to dismiss based on the defendant's jurisdictional minimum plea), the \$75,000 mentioned in the plea nevertheless served as an anchor and affected the judges' awards.

While the abovementioned examples demonstrate that experts are indeed susceptible to cognitive errors in their areas of expertise, they do not implicate expertise as *responsible* for engendering the errors. However, Arkes and Freedman (1984) demonstrated that expertise can be detrimental in some instances; that is, certain properties of expertise can lead experts to make mistakes where novices do not. In a study exploring the "paradox of interference," Arkes and Freedman (1984) showed that experts' ability to "go beyond the information given" (Bruner, 1957) led them to commit more errors than did novices in identifying new and old information. Arkes and Freedman (1984) presented baseball experts and non-experts with one of four stories about baseball. Two of the stories contained "synonymous targets" to be paired later

with “synonymous distractors,” which were synonyms of words used in the baseball story. The other two stories contained “inferential targets” to be paired later with “inferential distractors,” which were inferences that could be drawn from the story, although they themselves were not presented. All of the stories also contained “control targets” to be paired later with distractors that bore neither a synonymous nor an inferential relationship to the target. Participants had to read the assigned story, take a 45-question baseball quiz (used to categorize participants as experts and non-experts), and then rate sixteen sentences on a scale indicating to what extent the participants felt they had read the exact sentence in the baseball story (i.e., whether the sentence presented was “old” or “new”). Some of these sentences contained the control, synonymous and inferential distractors. Results demonstrated that when control distractors were included in the sentences, baseball experts showed superior recognition accuracy than did non-experts. However, when the sentences contained synonymous and inferential distractors, expert performance suffered. Specifically, when sentences contained synonymous distractors, experts were less likely than non-experts to judge the targets as “old” and the distractors as “new.” When sentences contained inferences that could be drawn from the story but were not actually presented, experts were more likely than non-experts to judge the distractor as “old.” The authors cleverly included as well an inference distractor whose inference would contradict what was presented in the story. In this case, expertise was beneficial; experts correctly identified the inferential distractor as “new” to a greater extent than did non-experts, since the experts recognized that the distractor was inconsistent with the events that took place in the story. Arkes and Freedman (1984) concluded that expertise is detrimental in some instances because experts can infer

information that was not actually presented; this otherwise beneficial ability to make inferences leads to poorer performance when the tasks calls for veridical recognition of presented information.

Related to Arkes and Freedman's (1984) findings is research on the relationship between confidence and expertise. Since experts are able to make inferences about presented information, they may infer that related stimuli were presented when they in fact were not (Arkes & Harkness, 1980). When such inferences are made, individuals may be unable to identify whether the information in question was actually presented or whether it was simply inferred from the information given; this is precisely what occurred with experts in Arkes and Freedman's (1984) study. Arkes and Freedman suggested that if an expert is later presented with information consistent with the information he or she had inferred previously, the presented information may be erroneously perceived as *additional* evidence supporting the inference the expert has already made. As evidence in favor of a particular conclusion accumulates, an inflated level of confidence may result.

In a study testing the effects of expertise on overconfidence, Bradley (1981) demonstrated that when participants were asked true/false questions, they preferred to guess rather than admit ignorance when the questions were in the topic area that the participant believed he had the greatest amount of expertise; this result is particularly disturbing in light of the finding that participants answered most of the questions in their area of expertise at or below chance frequency. Furthermore, participants were unwilling to admit to their uncertainty regarding the correctness of their answers; as the rank of area of expertise increased, the degree of uncertainty with regard to the correctness of the

answer decreased. More will be discussed about the implications of these results in later sections. While Bradley's (1981) demonstration of the effects of subjective expertise on overconfidence did not test the particular effect of expertise on confidence explicated by Arkes & Freedman (1984; i.e., that expertise inflates confidence due to an erroneous sense of accumulating evidence for a particular outcome), Bradley nevertheless established a connection between presumed expertise and confidence.

The research described in this chapter acknowledges the fact that expertise is beneficial in many instances, but it also brings to light the potential disadvantages of expertise in other cases. The following section explores a specific task in which expertise may be harmful or beneficial: predicting the performance of less-informed others.

CHAPTER 2

Hindsight Bias and Expertise: Is There a Curse of Knowledge?

The task of communicating information to less-informed others is a daily occurrence for professionals in many different lines of work. Professors must teach inexperienced students about calculus, nuclear fusion, the fall of the Holy Roman Empire, and countless other topics about which their students are otherwise relatively naïve. Physicians must explain to patients who have never learned about the intricacies of the nervous system why the pain in their lower back is exacerbated by running. Mechanics must relay to their mechanically unsophisticated customers why their car's engine is rattling, and financial planners must explain to their clients why it is good to have a diversified portfolio and safe to invest in government bonds. All of these situations require experts to relay information to non-experts in a way they will understand. As such, these situations require well-informed individuals to understand what less-informed individuals do and do not know and to correct for this knowledge discrepancy.

Anecdotally, it is clear that experts do not always succeed in their attempts to “dumb down” their communications enough for lay people to understand. Any student who has experienced complete confusion in a lecture in which he does not understand the

terminology the professor uses to explain other concepts recognizes some professors' inability to discount their special knowledge in the name of teaching. Similarly, any customer who leaves Jiffy Lube without knowing why his catalytic converter is making strange sounds, or what a catalytic converter even is, has been the victim of a mechanic's inability to put himself in the shoes of an individual who knows nothing about cars. In addition to anecdotal evidence, previous research has explored the problem of communication between more- and less-informed individuals. Nickerson (1999) noted that in order for individuals to communicate successfully to others, they first must gauge their own knowledge and then adjust this estimate for any special knowledge they have. In the case of a professor, he must assess his own level of knowledge, recognize that he has specialized knowledge that his students do not possess, and adjust downward to accommodate this difference. One potential problem with enacting this chain of events is that well-informed individuals may have difficulty identifying when they acquired the knowledge they currently possess (Nickerson, 1999). If a professor thinks he "always knew" how to take the first derivative of a function, he will fail to recognize that this information is something he must teach inexperienced students for them to understand how to solve math problems in his course. In a similar vein, individuals tend to overestimate the commonality of their knowledge (Nickerson, Baddeley, & Freeman, 1987); a professor who thinks taking the first derivative of a function is "common sense" will also fail to teach his students this important lesson. As a general finding, individuals mistakenly ascribe to less-informed others knowledge that they, themselves, possess (Nickerson, 1999), and this erroneous imputation of special knowledge can lead to breakdowns in communication.

In Hinds' (1999) discussion of the "curse of expertise," she identified reasons why well-informed individuals may find difficulty in adjusting their knowledge downward when communicating with less-informed individuals. First, Hinds implicated the *availability heuristic* (Tversky & Kahneman, 1973) as being responsible for experts' inability to discount their knowledge. The availability heuristic is a rule of thumb by which individuals estimate frequency or probability by the ease with which they can bring to mind instances of the stimulus in question. According to this explanation, experts can bring to mind easily instances in which they have dealt with the stimulus in an unproblematic way, leading the expert to overestimate the probability that others will be able to handle the stimulus in a similar trouble-free manner. Hinds (1999) also entertained the notion that *anchoring and adjustment* (Tversky & Kahneman, 1974) may contribute to the "curse of expertise." Here, when attempting to identify what others know, experts are presumed to "anchor" on their own level of knowledge and "adjust" downward to account for special knowledge they may possess. According to Tversky and Kahneman (1974) individuals typically fail to sufficiently move away from their original anchors; in the case of experts, then, Hinds (1999) suggested that it is the insufficient adjustment away from the anchor (one's own expert knowledge) that leads experts to overestimate what less-informed others know. A final reason Hinds (1999) gave for experts' failure to discount their special knowledge is experts' tendency to oversimplify tasks in their area of expertise. In the case of a calculus professor, such an oversimplification might lead the professor to say "...then you take the derivative of the function..." while omitting the steps involved in this procedure.

Research on the hindsight bias is intimately tied to the task of predicting what less-informed others know. First investigated by Fischhoff (1975), hindsight bias is the tendency to view past events as predictable or obvious after obtaining information about the outcome of the event. When individuals are provided with outcome information but must make decisions without taking into account the outcome information, they are essentially attempting to put themselves in the position of less-informed others. Regardless of whether an expert (e.g., Hinds, 1999) or non-expert (e.g., Nickerson et al. 1987) is called upon to make such a judgment, the individual must be able to discount his current knowledge state in order to respond appropriately to the question at hand. Previous research has demonstrated that experts and non-experts alike fall prey to the hindsight bias. For example, lay persons may view the results of an election as having been inevitable after learning who the victor is (e.g., Blank, Fischer, & Erdfelder, 2003; Leary, 1982), the events included in president Nixon's itinerary for his trip to China might seem foreseeable to the general public after he actually engaged in the various activities (Fischhoff & Beyth, 1975), and seasoned physicians may think the correct diagnosis is obvious after they find out about a patient's true ailment (Arkes, Wortmann, Saville, & Harkness, 1981).

The inability to discount outcome information has important implications for the law as well (e.g., Harley, 2007; Hastie, Schkade, & Payne, 1999; LaBine & LaBine, 1996), since legal verdicts by nature are decided after events have taken place and outcome information is available. Legal situations are an interesting area in which to examine hindsight bias in experts and non-experts, since both groups are called upon to make decisions in the courtroom: jurors are non-experts and judges are experts. Guthrie,

Rachlinski, and Wistrich (2001; 2007) examined the occurrence of hindsight bias in judges. Guthrie et al. (2001) gave 167 federal magistrate judges attending the 1999 Federal Judicial Center's Workshop for Magistrate Judges II in New Orleans a scenario regarding a prisoner who filed an action against his state's Director of the Department of Criminal Justice. The prisoner alleged that the prison provided him with negligent medical care, which the prisoner claimed was in violation of Section 1983. The district court dismissed the prisoner's claim because medical negligence does not violate Section 1983. In addition to dismissing the prisoner's claim, the district court also sanctioned the prisoner pursuant to Rule 11, which required that any future claims the prisoner wished to file be approved by the district's Chief Judge. This ruling was based on the fact the prisoner had filed similar medical negligence complaints in the past that were also dismissed, and he was therefore aware that medical negligence is not in violation of Section 1983. The prisoner proceeded to appeal the Rule 11 ruling. After reading the scenario, the judges in the experiment were assigned to one of three conditions. In the "lesser sanction" condition, the judges learned that the court of appeals ruled that the district court had abused its power under Rule 11 and ordered that a less severe version of Rule 11 be assigned to the prisoner. In the "affirmed" condition, the judges learned that the court of appeals upheld the district court's ruling of the imposition of Rule 11 on the prisoner. Finally, judges in the "vacated" condition learned that the court of appeals ruled that the district court had abused its power under Rule 11 and removed the prisoner's Rule 11 sanction. After receiving this outcome information, the judges were asked to assign probabilities to the three possible outcomes of the appeal described above. The results showed that the judges were greatly influenced by the ruling of the

court of appeals; the total percentage the judges assigned to the outcome that was “most likely to have occurred” was 172%. That is, the judges assigned much larger probabilities to the outcome they learned had actually occurred than did the judges who learned that a different outcome had occurred (note that if there were no effect of the outcome information, the total probability assigned to the “most likely to have occurred” category would have been 100%). Judges in this experiment clearly fell prey to hindsight bias.

Jurors, who are non-experts in legal decision making, also demonstrate the hindsight bias. Casper, Benedict, and Kelly (1988) examined whether or not jurors could disregard testimony presented during a court case that had been deemed inadmissible by a judge. Casper et al. (1988) gave mock jurors a scenario in which they read that police searched the apartment of a victim and either found drugs in the apartment or did not find drugs in the apartment. In a third group’s scenario, there was no mention of the search or the drugs. Since the search was conducted without a warrant, the mock jurors were instructed not to let the outcome of the search affect their award to the victim. Even though the mock jurors were told to disregard the evidence presented about the search, the average amount awarded to the victim by each group of jurors differed depending on the outcome of the illegal search. Jurors who learned that drugs were found in the victim’s apartment awarded the victim less money ($M = \$16,090$) than jurors learning that no drugs were found ($M = \$24,834$) or jurors who heard nothing about the search ($M = \$22,748$). The mock jurors in this experiment were unable to disregard the outcome information and act as if they had never learned about the search.

Berlin (2000) discussed an interesting medical malpractice lawsuit in which hindsight bias was demonstrated by both experts and non-experts in the same case. The lawsuit involved a radiologist who was sued for malpractice because he failed to detect a tumor on a patient's initial chest x-ray. Three years later, the patient returned to the physician because he had been experiencing chest pains and weight loss. A second x-ray was taken at this time and revealed a malignant tumor. The patient died 16 months after the second x-ray was taken. Expert radiologists were retained by both the plaintiff and the defendant to serve as expert witnesses at the trial. The expert radiologists testifying on behalf of the defense agreed that the tumor was not visible on the initial x-ray, while the experts testifying on behalf of the plaintiff argued that the tumor on the initial x-ray was clearly visible. Interestingly, the defense attorney included an exposition of hindsight bias in his line of questioning. While the plaintiff's experts claimed that hindsight bias was not responsible for their assertions because they reviewed the initial x-ray without first seeing the second, positive x-ray, the defense attorney argued that simply being asked to review x-rays for a malpractice case would have signaled to the radiologists that they were looking for a missed tumor. Thus, the defense attorney maintained that hindsight bias was the major determinant for the plaintiffs' expert witnesses' opinions that the tumor was visible on the initial x-ray. The jurors in the case found the defendant guilty of malpractice. This case is an example of experts (the plaintiffs' expert witnesses) and non-experts (the jurors) falling prey to the hindsight bias; the former demonstrated the bias in their determination that the tumor was clear on the initial x-ray, and the latter may have been influenced by the knowledge that the patient was about to die as a result of the tumor. Jurors in medical malpractice cases are instructed to make decisions based only

on the physician's standard of care at the time of incident and not on the resulting events, but jurors are often influenced by whether or not damages have occurred (LaBine & LaBine, 1996).

Like the radiologists described above, expert physicians have also been shown to demonstrate the hindsight bias. Dawson, Arkes, Siciliano, Blinkhorn, Lakshmanan, and Petrelli (1988) examined the occurrence of the hindsight bias in physicians' predictions of medical diagnoses. During a clinicopathological conference, an instructional session in which a discussant reports on a medical case and selects the diagnosis he deems most likely, Dawson et al. asked half of the physicians in the audience to predict the probability that each of 5 possible diagnoses was correct. After these predictions were made, the audience members learned the patient's actual cause of death, and the other half of the physicians in the audience were asked to report the probabilities that they would have assigned to each of the 5 diagnoses before hearing the cause of death. Dawson et al. found that physicians who were afforded the information about the patient's actual cause of death assigned a much higher probability to the true diagnosis than did participants who assigned probabilities before learning the outcome information. While this study showed that experienced physicians were susceptible to the hindsight bias, the most senior physicians did not demonstrate the bias on the most difficult cases; these most experienced doctors could appreciate the rarity of the co-occurrence of two symptoms as indicative of the ailment responsible for the patient's death.

Although the abovementioned examples demonstrate that even experts can fall prey to hindsight bias, some recent research suggests that expertise can protect individuals from hindsight bias. In a virtual-reality batting task, Gray, Beilock, and Carr

(2007) examined how well expert and novice baseball players would be able to predict the outcomes of their bat swings. Using virtual-reality technology, Gray et al. were able to pause the simulation at the exact moment the bat hit the ball. At this point, participants were asked to predict where on the field the ball would land by clicking on a yellow cone on the screen and moving it to the location on the field they thought accurately reflected where the ball would land. If they thought they hit a homerun or foul ball, participants placed the yellow cone where they thought the ball would cross the fence or exit the playing field. After making these predictions, participants were given accurate feedback as to where the ball would have actually landed. After participants received outcome feedback, the virtual screen went blank and reappeared after 20 seconds. At this point participants had to mark again where they thought the ball would have landed without taking into account the outcome feedback. Results showed that the discrepancy between the placement of the yellow cone before and after participants received outcome feedback was greater for novice players than it was for expert players. The authors included an additional analysis that tracked hindsight bias as a function of expertise and current performance and found that it was when experts were doing well at the task that hindsight bias was reduced. While the results of this study are compelling, it is difficult to confidently generalize the results of this research to other hindsight studies for two reasons. First, this study tested a very specific skill (swinging a baseball bat and predicting the result of the swing) that may be idiosyncratic with regard to its relationship to hindsight bias. For example, expert baseball players have an enormous amount of veridical feedback regarding the result of each swing they take. In this sense, baseball experts asked to predict the outcome of a swing are like “super experts” as compared to

experts tested in other domains in which feedback is much more imperfect and infrequent. Shanteau (1992) commented on this distinction, noting that experts whose domain of expertise involves repetitive tasks that produce reliable feedback tend to make more accurate predictions than do experts in tasks that involve unstable and unreliable stimuli. Second, the evidence for visual or perceptual hindsight bias has been mixed. Winman, Juslin, and Bjorkman (1998) found a reverse hindsight bias in a perceptual task, while Harley, Carlsen and Loftus (2004) found a positive hindsight bias in a perceptual task (see also Bernstein & Harley, 2007).

Recall the abovementioned study conducted by Guthrie et al. (2001) in which judges displayed hindsight bias in an appeals case. While the authors demonstrated that even magistrate judges fall prey to the hindsight bias in their area of expertise, a second study failed to demonstrate the bias in judges. Guthrie et al. (2007) discussed an experiment in which judges were presented with a case regarding probable cause. The judges read a scenario describing a police officer who was patrolling a parking lot outside of a rock concert when he noticed a concert-goer nervously looking through the trunk of a BMW. A half hour later the officer became aware that the window to the car was rolled down and approached the car in order to close the window. Upon approaching the car, the officer smelled something he thought was likely burnt methamphetamine, and he proceeded to look inside the car only to find Visine, a map, and some empty beer cans. After reading this first part of the scenario, the judges were randomly assigned to either a foresight or hindsight group. Foresight judges read that the officer called in a request for a telephonic search warrant, since he believed there was probable cause to search the trunk of the car. The foresight judges were then asked whether or not they thought there

was probable cause to grant the telephonic warrant based on the scenario. Judges in the hindsight group were told that the officer searched the trunk of the car without a warrant and found ten pounds of methamphetamine and other drug paraphernalia, as well as a gun that was used to murder a drug dealer earlier that day. In the hindsight scenario, the officer proceeded to arrest the car owner, who was later prosecuted even though the defense attorney argued in court that the evidence found in his client's trunk was inadmissible on grounds that the officer did not have probable cause to search the trunk. Hindsight judges were then asked whether they would admit the evidence from the warrantless search. The percentage of judges in the foresight and hindsight groups who thought there was probable cause to grant the warrant (23.9%) and who would have admitted the evidence procured from the search (27.7%), respectively, were not significantly different. That is, unlike the results of the other study by Guthrie et al. (2001) in which judges did demonstrate the hindsight bias, judges in this probable cause case did not fall prey to the hindsight bias. Guthrie et al. (2007) argued that the strict, rule-based nature of the probable cause problem in jurisprudence may have aided the judges in their decision-making process, shielding them from hindsight bias. Indeed, Guthrie et al. (2007) noted that many judgments and decisions judges make on a daily basis are made intuitively as opposed to analytically. The addition of strict rules to the decision-making process may shift the judge's decision strategy from intuitive to analytical, thereby reducing the hindsight bias. Previous researchers (e.g., Arkes, Faust, Guilmette, & Hart, 1988; Koriatic, Lichtenstein, & Fischhoff, 1980) have also demonstrated that analytical thinking (e.g., "consider the opposite") can help to reduce the hindsight bias.

The jury seems to still be out on the exact relationship between hindsight bias and expertise. In the current series of experiments, I argue that expertise exacerbates hindsight bias because experts *feel like* they know more in their areas of expertise than they actually do. Demonstrating that hindsight bias increases with expertise is interesting because it is counterintuitive. It would seem that those who are more knowledgeable in a particular area should be less vulnerable to biases in the area (e.g., Fischhoff, 1982; Smith & Kida, 1991). In fact, in their meta-analysis of previous hindsight bias studies, Christensen-Szalanski and Willham (1991) found that one moderator of hindsight bias is “expertise” or “familiarity” with a particular task. The authors coded 128 experiments as either “familiar” or “not familiar” based on the types of questions that were asked in the experiments and who the participants were. For example, the abovementioned study by Dawson et al. (1988) was coded as “familiar,” since physicians were asked about medical diagnoses. Christensen-Szalanski and Willham found in their analysis that there was an effect of familiarity on hindsight bias: the effect of hindsight bias was smaller for more familiar tasks. This result supports the notion that those who know more in a particular area should be less susceptible to biases in this area but contradicts the hypothesis put forth in the current series of experiments (i.e., that expertise exacerbates hindsight bias).

If a meta-analysis of hindsight bias has already demonstrated that expertise is negatively related to hindsight bias, why would I bother conducting a study on this topic, let alone propose the opposite? While Christensen-Szalanski and Willham (1991) found that familiarity moderated the hindsight bias, their definition of “familiarity” is quite different from the type of familiarity or expertise discussed in the current series of experiments. Christensen-Szalanski and Willham (1991) categorized existing hindsight

experiments as “familiar” if the participants were experts in the subject matter presented, but they also categorized as “familiar” studies in which the “subjects had actually experienced the outcome or task (e.g., watched a team lose a basketball game”; p. 151). This latter part of the definition is unrelated to the type expertise explored in the current experiments, so it is unclear how the results of Christensen-Szalanski and Willham’s (1991) meta-analysis should be interpreted for the purposes of this paper. In a more recent meta-analysis of 95 hindsight bias studies, Guilbault, Bryant, Brockway, and Posavac (2004) found results dissimilar to those previously found by Christensen-Szalanski and Willham (1991). Guilbault et al. (2004) did not observe a relationship between expertise and hindsight bias, a difference the authors attributed to their different categorization of “familiarity” or “expertise.” Guilbault et al. (2004) noted that Christensen-Szalanski and Willham’s (1991) classification of studies in which participants were familiar with the task was too broad and subjective, so Guilbault et al. (2004) limited studies categorized as “expert” in their meta-analysis to those containing participants who were actually experts in the area being tested (e.g., Arkes et al., 1981). This change in the definition of “expertise” produced a different result: hindsight bias was not moderated by expertise. The definition of expertise Guilbault et al. (2004) used in their meta-analysis more closely resembles the type of expertise referred to in the current series of experiments. While Christensen-Szalanski and Willham (1991) found a negative relationship between expertise and hindsight bias, and Guilbault et al. (2004) found no relationship between expertise and hindsight bias, it remains the case that there are numerous examples in previous literature in which experts do fall prey to the

hindsight bias. In the current series of experiment I attempt to reconcile these conflicting findings.

While most of the experiments described in this section demonstrate that experts either do or do not demonstrate the hindsight bias in their areas of expertise, they do not discuss how differences in degrees of expertise can account for differences in the degree of hindsight bias exhibited. Two exceptions are a) the findings from Dawson et al.'s (1988) study, which showed that although experienced physicians were susceptible to the hindsight bias, the most senior physicians did not demonstrate the bias for the most difficult cases, and b) the findings from Gray et al.'s (2007) study, which showed that baseball experts displayed a smaller hindsight bias in a perceptual task than did non-expert players. Only a few additional studies have looked at differences in hindsight bias as a function of expertise and have found conflicting evidence (e.g., Pohl, 1992; Pohl, Schwarz, Sczesny, & Stahlberg, 2003; Smith & Kida, 1991). In the current series of experiments, I test directly how differences in levels of expertise affect the occurrence of hindsight bias and conclude that expertise generally exacerbates hindsight bias (but, see Experiment 2).

Why might hindsight bias be more prevalent in experts than in non-experts? Such an effect may occur if people judge what they know based on the *domain* in which the question lies rather than judging based on the specific question being asked. In other words, an individual's perception of what he knows in a particular domain likely exceeds what one *actually* knows in that domain. This general effect is often expressed as a measure of overconfidence, and it has not only been observed in experts (e.g., Camerer & Johnson, 1997; Koheler, Brenner, & Griffin, 2002), but it also has been shown to vary

with expertise (e.g., Bradley, 1981; McKenzie, Liersch, & Yaniv, 2008; Yates, McDaniel, & Brown, 1991). For example, Dawson, Connors, Speroff, Kemka, Shaw, & Arkes (1993) demonstrated that physicians' confidence in their estimates of values pertinent to right-heart catheterization was unrelated to their accuracy of these estimates; the physicians were overconfident in all instances. In fact, the more experience the physicians had, the more confident they were, but experience afforded no increase in accuracy. Thus, more experienced physicians displayed greater overconfidence than did those who were less experienced. Bradley (1981) also found a relationship between confidence and expertise. He demonstrated that the reluctance to admit ignorance increased as subjective expertise increased even though accuracy in one's area of presumed expertise was below chance. Participants also were unwilling to admit to their uncertainty regarding the correctness of their answers; as the rank of area of expertise increased, the degree of uncertainty with regard to the correctness of the answer decreased. Inasmuch as certainty in the correctness of one's answer can be expressed as the degree to which one feels like he "knows the answer," the effects of expertise on hindsight bias ("I knew it all along") should mirror the effects of expertise on overconfidence. Indeed, Bradley (1981) noted that participants showed "a reluctance to admit ignorance when one should, the reluctance being most pronounced in one's area of greatest expertise" (p. 84). It is precisely this reluctance to "admit ignorance" that may contribute to an increase in hindsight bias as expertise increases; failure to admit ignorance is comparable to an individual's feeling that he "knew it all along."

It is interesting to note that in addition to individuals' feelings that their expertise extends far beyond the actual boundaries of one's expertise, individuals often think that

others' areas of expertise are more inclusive than they actually are. Some common interactions between experts and non-experts may inform the discrepancy between what experts know and what experts think they know. My friends often ask me to "analyze them," or they ask "Do you know what I'm thinking right now?," because they know that I am a graduate student in psychology. Non-experts may not realize that there are in fact sub-disciplines that contribute to expertise in some areas within a domain without contributing to expertise in other areas within the same domain. If an individual has a cardiologist in his family, for example, the individual may be tempted to ask the cardiologist about a strange growth on his back simply because the family member is a physician. Again, the individual may overgeneralize the extent of the cardiologist's expertise, assuming that he knows everything about medicine because he has a medical degree. In some sense, however, the cardiologist's family member is not completely unjustified in asking the cardiologist about a medical condition, since the cardiologist most certainly knows more about medicine than does the family member, who has no medical training. My mother has long gotten irritated by her friends' comments that "you're a teacher, you should know that" when she does not know the answer to a particular question. My mother was a kindergarten teacher and a science specialist for 35 years, and although she is quite intelligent, she should not be expected to know everything about U.S. history simply because she is a teacher. This general failure to recognize that expertise in a particular sub-topic of a domain does not denote expertise in the entire domain can contribute to discrepancies between what one knows and what one thinks he knows as well as what one knows and what others think he knows. The examples mentioned above highlight an important aspect of the difficulty in identifying

expertise: is expertise a relative concept or an absolute concept? A discussion of this issue appears in Chapter 5.

The current hypothesis that greater expertise leads to greater hindsight bias stems from previous research demonstrating that greater confidence in one's answers is associated with greater hindsight bias (Werth & Strack, 2003), and that familiarity with a topic leads to greater hindsight bias (Marks & Arkes, 2009). Recall the discussion above regarding the effects of expertise on overconfidence. Bradley (1981) showed that overconfidence increased as subjective expertise increased. Werth and Strack (2003) demonstrated more recently that greater subjective feelings of confidence can lead to an increased "knew-it-all-along" effect. Werth and Strack (2003) gave participants 40 extremely difficult questions (i.e., they only chose questions that no participants in a pretest answered correctly) requiring numerical estimates as answers. Participants were also given one of two answers for each question: a low estimate and a high estimate of the actual numerical value. The participants were instructed to answer the questions as they would have had they not been given the answer, and they had to give confidence ratings indicating how confident they felt about each answer. Werth and Stack (2003) found greater assimilation to the provided estimates when participants had experienced more confidence in generating their answers. The authors argued that this effect is the result of participants experiencing a sense of familiarity or a sense that "the question seems so familiar" that leads to confidence and, in turn, hindsight bias.

While Werth and Strack (2003) contended that greater confidence leading to greater hindsight bias resulted from participants feeling like the information was more familiar, they did not directly test this assumption. In a second study, Werth and Strack

(2003) manipulated perceptual fluency in an attempt to show that greater fluency similarly would lead to feelings that the question “seems familiar to me,” but this second experiment still did not directly test the familiarity assumption. In a series of eight experiments, Marks and Arkes (2009) explored the specific idea that familiarity leads to the hindsight bias. The authors showed that participants were unable to discount information they read about the familiar Revolutionary War when taking a subsequent true/false quiz, but they were able to disregard information they read about the more obscure War of 1812 during a subsequent quiz. Marks and Arkes (2009) showed that *source confusion* was responsible for the difference in participants’ ability to disregard essay information. Source confusion refers to the failure to correctly identify the source of one’s current knowledge (Wilson & Brekke, 1994). Marks and Arkes (Experiment 3) found that familiar information is more susceptible to source confusion than is unfamiliar information, and it is this difference in source confusion that leads to participants’ differential ability to discount these different types of information.

In the current series of experiments I further explore the mechanisms underlying hindsight bias by establishing a connection between expertise and hindsight bias. By definition, experts should be more familiar than non-experts with information dealing with their area of expertise. If hindsight bias is more likely to occur with more familiar information (Marks & Arkes, 2009), then expertise, like familiarity, should be a determinant of hindsight bias as well. As mentioned previously, some studies have already explored the connection between hindsight bias and expertise (e.g., Dawson et al., 1988; Christensen-Szalanski & Willham, 1991; Guilbault et al, 2004), but the findings thus far have been inconsistent and inconclusive. In the current study, I present a

process model of the effect of expertise on hindsight bias, contending that the reason expertise leads to hindsight bias is via an increased “feeling of knowing” in one’s area of expertise. Furthermore, I demonstrate that there may be an inverted U-shaped relationship between expertise and hindsight bias, such that hindsight bias is lower for novices and “super experts” but higher for “semi-experts” (McKenzie et al., 2008).

CHAPTER 3

Experiment 1: The Effects of Source Confusion on Hindsight Bias in Experts and Novices

Marks and Arkes (2009) used a novel methodology in their demonstration of hindsight bias. In typical hindsight bias experiments, participants in a hindsight group are given both a question and the answer to the question and are asked to give the probability that they would be able to correctly answer the question had they not been given the answer. The average probability given by participants in the hindsight group is then compared to the proportion of participants in a foresight group who answered the question correctly. The discrepancy between these two averages is the measure of hindsight bias. Rather than assessing hindsight bias on a question by question basis, Marks and Arkes (2009) asked participants to recreate a prior, naïve state of knowledge. Participants were asked to read an essay about the Revolutionary War or the War of 1812 and subsequently were asked to answer a 20-question true/false quiz as if they had not read the essay. This methodology required participants to assess what they knew about either of these wars before reading the essay during the experiment. Participants reading about the Revolutionary War had difficulty recreating their prior knowledge state, as the knowledge they possessed after reading the essay included both prior knowledge and knowledge acquired while reading the essay during the experiment. Since participants

came to the experiment with some prior knowledge about the Revolutionary War, they were unable to successfully separate what they knew before reading the essay from what they learned while reading the essay; that is, participants reading about the Revolutionary War experienced source confusion (Wilson & Brekke, 1994). Indeed, when the authors cued participants to separate new and old Revolutionary War information while they read the essay (Experiment 5), participants exhibited significantly less hindsight bias.

Participants reading about the War of 1812, on the other hand, had much less difficulty segregating new and old information, since they came to the experiment knowing very little about the War of 1812. For the War of 1812 participants, recreating a prior knowledge state was easier since there was less source confusion, and consequently these participants demonstrated less hindsight bias than did participants reading about the Revolutionary War.

The first experiment in the current series of experiments borrows the methodology used by Marks and Arkes (2009) to explore the effects of expertise on hindsight bias. Rather than manipulating the familiarity of the information presented to participants (Revolutionary War vs. War of 1812), I held constant the presented information and assessed how participants dealt with the same information as their level of expertise varied. Specifically, participants read an essay about the 2006 major league baseball (MLB) most valuable player (MVP) race or about owning a dog and subsequently answered a 20-question true/false quiz about the essay; some participants were asked to answer the questions to the best of their ability while others were asked to disregard the information from the essay. Participants were then asked to answer a 33-question trivia-type quiz about baseball rules (adapted from Arkes & Freedman, 1984),

and participants' score on this second quiz was used as a measure of expertise. I predicted that as baseball expertise increased, the inability to disregard the essay information would increase as well. That is, hindsight bias would increase as expertise increased.

Method

Participants. Two-hundred sixty-two undergraduates enrolled in Psychology 100 at Ohio State University participated in this study. Each student received course credit for his or her participation.

Materials. The essays were presented individually to each student on a computer using the MediaLab program. After reading the appropriate essay, each student answered the 20-question true/false quiz about the 2006 MVP race on the computer as well (see Appendix A for essays and quiz). The MVP race essay contained all of the information needed to answer the subsequent 20-question true/false quiz and each question in the 20-question quiz came directly from sentences in the MVP race essay. An attempt was made to include a quiz question pertaining to each fact or topic presented in the essay so that the number of questions taken from each section of the essay (i.e., beginning, middle, and end) was approximately even. The quiz questions were presented in a fixed order, corresponding to the text of the essay, to ensure that the exposure time between the reading of a particular sentence and the answering of the quiz question pertaining to that sentence was held relatively constant between participants.

In addition to the essays and the true/false quiz, participants were also presented with 33 trivia-type questions about baseball rules (see Appendix B for quiz). These

questions were also presented on the computer, and participants had to type the answers on the computer in the space provided.

Procedure. Participants were randomly assigned to one of three conditions and were asked to complete a true/false quiz immediately after reading a particular essay. Participants in condition 1 ($n = 83$) were given instructions on the computer to read an essay about the MVP race and answer to the best of their ability a subsequent 20-question true/false quiz about the information they just read. This group was called the “do-your-best” (DYB) group, and its purpose was to check if reading the essay positively impacted participants’ quiz performance. Participants in condition 2 ($n = 96$), the “discount” condition, were given instructions on the computer to read an essay about the 2006 MVP race and answer a subsequent 20-question true/false quiz as if they were a typical Ohio State University (OSU) student who had not just read the essay. Specifically, participants in the “discount” group were given the following instructions on the computer:

“We realize that you have just read an essay about the 2006 MVP race, but we would like you to take the following 20-question true/false quiz as if you had not just read the essay. In other words, suppose you were a typical OSU student who had not read the essay. Without that knowledge, how would a typical OSU student have answered the following 20 questions?”

Finally, participants in condition 3 ($n = 83$) were given instructions on the computer to read an essay about owning a dog and answer a subsequent 20-question true/false quiz about the 2006 MVP race. This group was called the “dogs” group and it was employed to determine how much prior knowledge participants had about the 2006 MVP race. For each condition, participants were allowed to read the essay at their own pace, and they

were instructed to move on to the questions once they finished reading the essay. The true/false questions then appeared one at a time.

Once participants answered the 20 true/false questions, they were told that they would move on to the second part of the experiment, which consisted of a 33-question trivia-type quiz about baseball rules. Participants were instructed to type the correct answer to the question in the space provided. They were told that if they could not answer the question, they should try to make an educated guess or type “I don’t know” if they could not provide an educated guess.

Results. Participants’ scores on the 33-question trivia quiz were used as the measure of baseball expertise (called the “expertise quiz” from this point on). Scores on this quiz ranged from 0 questions correct to 31 questions correct ($M = 13.73$, $sd = 8.66$; see Figure 1). As a measure of hindsight bias, I examined participants’ ability to discount the essay information; specifically, I looked at the difference between the “dogs” scores and the “discount” scores as expertise increased. The rationale for using the dogs scores as the basis for comparison lies in the instructions given to “discount” participants during the experiment. Recall that “discount” participants were asked to answer the quiz questions “as if they were a typical OSU student who had not read the essay.” Therefore, participants in the discount group should have performed like participants who truly had not read the essay, namely the “dogs” participants.

In order to analyze the relationship between the “dogs” and “discount” groups as a function of expertise, I conducted a regression analysis in which I predicted the number correct on the MVP race true/false quiz from group membership (DYB, discount, dogs), expertise quiz score, and the interaction between expertise and membership in the DYB

and discount groups. In order to make membership in the dogs group the basis of comparison, I dummy coded group membership so that DYB and discount scores were relative to dogs scores. The regression analysis revealed that both interaction terms were significant ($b_{\text{DYB}*\text{expertise}} = .224$; $b_{\text{discount}*\text{expertise}} = .127$, both p 's < .01), indicating that group membership interacted with expertise in predicting scores on the 20-question MVP quiz.

The regression weights presented in Table 1 can be best described using representative participants from each group. The value of the constant indicates that a participant in the dogs group (the reference group for this analysis) who answered 0 questions correctly on the expertise quiz would be expected to answer 9.915 questions correctly on the 20-question true/false quiz (MVP quiz). Since the MVP quiz was comprised of 20 true/false questions, participants would be expected to answer 10 questions correctly by chance (i.e., if they guessed on each question). The predicted MVP quiz score for a dogs participant who was completely naïve with regard to baseball rules is approximately at chance level. A participant in the discount group who scored a 0 on the expertise quiz, however, is expected to score 1.225 points higher on the MVP quiz than would a dogs participant, and a DYB participant scoring a 0 on the expertise quiz would be expected to score .57 points higher on the MVP than would a dogs participant. As participants answer more questions correctly on the expertise quiz, the significant interaction terms lead to different predictions for participants' MVP quiz scores depending on their group membership. A participant in the dogs group who answers 10 questions correctly on the expertise quiz would be expected to answer 10.25 questions correctly on the MVP quiz, while a discount participant answering the same

number of questions correctly on the expertise quiz would be expected to answer 12.74 questions correctly on the MVP quiz (a difference of 2.49 questions). As the number of questions answered correctly on the expertise quiz increases (i.e., as expertise increases), the difference between the predicted values of dogs and discount participants increases as well. For example, participants in the dogs and discount groups answering 25 questions correctly on the expertise quiz are now expected to answer 10.74 and 15.14 questions correctly, respectively, on the MVP quiz. While the difference between dogs and discount participants' MVP scores was 2.49 when the expertise score was 10, the difference between the MVP scores when the expertise score is 25 is 4.4. This increase in the difference between MVP scores of dogs and discount participants as expertise increases is evidence for the prediction that the inability to discount (i.e., the inability for discount participants to act like dogs participants who never read the MVP essay) increases as expertise increases (see Figure 2).

Another way to interpret the same results is to consider how scores on the MVP quiz changed for participants in each group as a function of expertise quiz score. Participants in the dogs group were only afforded 1 more correct answer on the MVP quiz for answering 30 more questions correctly on the expertise quiz. That is, a huge difference in one's expertise quiz score resulted in only a very small difference in one's MVP quiz score if participants were in the dogs group. However, for participants in the discount group, answering only 6 more questions correctly on the expertise quiz afforded them 1 more correct answer on the MVP quiz. Finally, participants in the DYB group only had to answer 3 more questions correctly on the expertise quiz to have a predicted 1 question increase on the MVP quiz.

The regression results suggest, therefore, that while greater expertise did not help participants perform better on the MVP quiz if they had not read the essay (dogs group), greater expertise did lead DYB and discount participants, who did read the MVP essay, to score higher on the MVP quiz. However, recall that discount participants were not supposed to take the essay information into account when taking the MVP quiz. Thus, it seems that greater expertise hurt discount participants, since they became less able to “unlearn” the essay information as their level of expertise increased. Additionally, these results suggest that experts did not necessarily know more about the 2006 MVP race (i.e., expertise barely affected dogs participants’ scores), but upon reading the essay, the experts felt that they would have known more about the MVP race without reading the essay than they actually would have.

Discussion. The results of Experiment 1 demonstrate that the ability to discount is negatively associated with expertise; as baseball expertise increased so did the inability to discount the essay information about the 2006 MVP race. Although all of the participants read the same essay, the discounting task proved more difficult for those who possessed greater baseball expertise. Based on the findings of Marks and Arkes (2009), the results of Experiment 1 suggest that as the level of participants’ baseball expertise increased, participants found it more difficult to separate what they knew before reading the essay from what they learned while reading the essay. Similar to Marks and Arkes’ (2009) findings that hindsight bias was greater with familiar than with obscure information, Experiment 1 demonstrated that hindsight bias increases as expertise increases.

The results of Experiment 1 are consistent with the source confusion account Marks and Arkes (2009) used to explain their findings. Marks and Arkes (2009) showed that participants were unable to discount Revolutionary War information but were able to discount War of 1812 information. The authors argued that since participants were already familiar with the people and events involved in the Revolutionary War, they experienced source confusion when they read an essay during the experiment that dealt with this familiar information; that is, participants were confused as to whether the source of their current knowledge was the essay they read during the experiment or the knowledge they possessed prior to their participation in the experiment. In this sense, participants in the current experiment who answered many questions correctly on the expertise quiz were analogous to participants in Marks and Arkes (2009) study who read about the Revolutionary War. Baseball experts in Experiment 1 likely experienced confusion regarding the source of their current knowledge about the 2006 MVP race after reading the essay, since they were already familiar with the people and events surrounding the race. When experts were asked to act as if they did not read the essay, like Revolutionary War participants in Marks and Arkes' (2009) study, the experts were unable to identify which information was new and which information was old, leading to poor performance on the discounting task. Non-experts, other the other hand, were analogous to Marks and Arkes' (2009) participants who read about the War of 1812. Both sets of participants are presumed to have known very little about the events and people mentioned in the essay before participating in the experiment. Thus, when these participants were asked to discount the essay information, they were able to recognize that any knowledge they currently possess about the event in question must have come

from the essay; this lack of source confusion enabled participants to successfully discount the essay information.

There are two caveats to address in comparing the results of Experiment 1 to Marks and Arkes' (2009) study. First, there was likely more variability in the independent variable (expertise) in the current study than there was in Marks and Arkes' independent variable (which war participants read about). Expertise in the current experiment was a continuous variable, and participants could have scored anywhere between 0 and 33 on the expertise quiz. The independent variable in Marks and Arkes' study, however, was dichotomous; participants either read about the Revolutionary War or about the War of 1812. While participants in Marks and Arkes' study could have varied in the amount of knowledge they possessed about each war, this variability was not directly measured like it was in the current experiment. Therefore, the War of 1812 participants in Marks and Arkes' study are not completely analogous to "non-experts" in the current study, since there is not a cut-off that separates experts from non-experts in Experiment 1. It is clear that participants answering 0 questions correctly on the expertise quiz were completely naïve, but it is unclear how to categorize the rest of the participants in terms of their expertise. All that can be concluded from the current results is that hindsight bias increased as expertise increased; nothing can be said about exactly how expert the participants were. In fact, results from Experiment 2 (reported in Chapter 4 below) suggest that even participants scoring highest on the expertise quiz may not be considered "true" experts. As alluded to in Chapter 2, expertise may be best considered in relative terms, rather than in absolute terms. While expertise is often referred to as a dichotomy (e.g., experts vs. novices or experts vs. non-experts) in the literature, it may be

best to consider expertise as a continuum, as was done in Experiment 1 by using expertise quiz scores as a measure of expertise.

A second stipulation in the interpretation of the current results is that source confusion was not directly measured in Experiment 1 like it was in Marks and Arkes' (2009) study (Experiment 3). In a direct test of source confusion, Marks and Arkes found that participants could discriminate better between information presented and not-presented in the War of 1812 essay than in the Revolutionary War essay, indicating greater source confusion with Revolutionary War information. I would expect participants in the current experiment who scored lower on the expertise quiz to be able to discriminate between presented and not-presented information better than those who scored higher on the quiz, indicating greater source confusion for those who were more expert. This direct test of source confusion could be conducted in a future study.

It is difficult to compare these results to those found in hindsight bias studies other than Marks and Arkes' (2009) because a novel methodology was used. Nevertheless, the results of Experiment 1 support the findings of Werth and Strack (2003), who demonstrated that participants who expressed greater confidence in their answers showed a greater "knew-it-all-along" effect. Previous literature (e.g., Bradley, 1981; Dawson et al., 1993) has demonstrated that confidence increases with expertise. Thus, if participants in Experiment 1 were more confident the more expert they were, the results presented here would corroborate Werth and Strack's (2003) findings. In fact, the current results provide a stronger test of Werth and Strack's assertion that familiarity is responsible for the knew-it-all-along effect, since Werth and Strack (2003) did not directly test their familiarity hypothesis.

CHAPTER 4

Experiment 2: Are “Super Experts” Resistant to the Hindsight Bias?

As previously noted, the methodology used by Marks and Arkes (2009) and in Experiment 1 above is different from the methodology typically used in hindsight bias studies. Traditionally, participants are randomly selected to either answer questions (foresight group) or, after seeing both a question and the correct answer to the question, to give the probability that they would have been able to answer the questions correctly had they not been provided with the answers (hindsight group). The hindsight probabilities are then compared to the proportion of foresight participants who answered the question correctly. In Experiment 2 I conducted a traditional hindsight bias study to test the hypothesis that hindsight bias is exacerbated in one’s area of expertise. Demonstrating the effects of expertise on hindsight bias using the traditional paradigm will allow for more direct comparisons between the current experiment and previous hindsight bias studies.

True to the definition of expertise, the proportion of baseball “experts” to “non-experts” in the Psychology 100 subject pool was rather small. In fact, approximately 14% of participants tested were considered experts by the standards I identified for this experiment (the expert/novice selection procedure is explained in the “Methods” section

below). Therefore, the sample sizes in the foresight and hindsight expert groups were too small to use confidently in statistical analyses. In order to remedy this problem, I recruited players from the Ohio State University varsity baseball team to complete the experiment; I could be sure that 100% of these participants would be considered experts by any standard. After collecting the data from the baseball players, I discovered that these experts were even more expert than the participants I considered experts in the Psychology 100 subject pool, and this difference in expertise, even among experts, proved to be meaningful in terms of the dependent variable. While this extremely interesting discovery was not anticipated, it has added an important facet to the exploration of the relationship between expertise and hindsight bias and will be discussed throughout this section.

Method

Participants. One-hundred twenty-seven undergraduates enrolled in Psychology 100 at The Ohio State University participated in this study. Each student received course credit for his or her participation. After imposing the constraints on level of expertise, 101 participants remained in the analysis. In addition to participants from the Psychology 100 subject pool, 32 Ohio State University varsity baseball players participated in this study.

Materials. Participants saw a series of 20 questions dealing with current events in baseball (see Appendix C for questions). The questions were presented individually to each student on a computer using the MediaLab program. Participants were instructed to type their answers onto the computer in the space provided. The baseball players

completed a paper-and-pencil version of the questionnaire that the baseball coach administered to the players in the locker room.

In addition to the 20 foresight/hindsight questions, participants were also presented with the same 33-question baseball quiz used in Experiment 1 as a measure of expertise. These questions were also presented on the computer, and participants had to type the answers on the computer in the space provided. The baseball players completed an abridged paper-and-pencil version of the expertise quiz. Since the baseball players gave their time with no compensation, I wanted to make the task as minimally effortful as possible for them. Therefore, I analyzed the expertise quiz scores for the Psychology 100 participants and identified the 9 most difficult questions (i.e., the questions the least Psychology 100 students answers correctly) with the assumption that the members of the varsity baseball team would be able to answer all of the easier questions. The baseball players answered only these 9 most difficult questions, and only 24 of the 32 baseball players who completed the foresight/hindsight portion of the experiment completed the 9-question expertise quiz.

Procedure. Participants were randomly assigned to the foresight group or the hindsight group. Participants in the foresight group saw the questions on the computer screen (or on individual pieces of paper in the case of the baseball players) and were instructed to type (write) the answer to each question in the space provided. Participants in the hindsight group saw each question, and they also saw the answer to each question in parentheses after the question. For example, participants in the hindsight group saw the following question:

“Who was named the 2005 American League Most Valuable Player (MVP)?
(Alex Rodriguez)”

Participants in the hindsight group were instructed to rate on a scale of 0-100 the probability that they would have answered the question correctly had we not provided them with the correct answer.

Once participants answered the 20 foresight/hindsight questions, they were told that they would move on to the second part of the experiment which consisted of a 33-question (9-question) trivia-type quiz about baseball. Participants were instructed to type (write) the correct answer to the question in the space provided. They were told that if they could not answer the question, they should try to make an educated guess or type (write) “I don’t know” if they could not provide an educated guess.

Expert/Novice Selection Process. Psychology 100 participants’ scores on the 33-question baseball expertise quiz ranged from 1 to 29 ($M = 13.02$, $sd = 8.63$; see Figure 3). I implemented the following standards to identify experts and non-experts within the sample. I considered experts to be those who scored between 25 and 29 on the 33-question quiz. Though somewhat arbitrary, I used the highest score (29) and subtracted 5 from this score as the lowest a participant could have scored while still being considered an expert. I reasoned that participants who answered 5 more questions incorrectly than those who scored the highest could no longer be considered experts. Non-experts were participants who scored between 1 and 17. Seventeen is the score that falls at the top of the middle third of the distribution of scores for this sample. Note that I did not consider non-experts to be those who scored just below the experts (24), since I do not think that a single question can separate experts from non-experts. As such, I eliminated data from

participants who scored between 18 and 24 on the quiz (26 participants). As mentioned earlier, expertise may be best thought of as a continuum rather than a dichotomy; while I dichotomized expertise by designating participants as “experts” and “non-experts” in the current experiment, I did so only after removing from the distribution participants who scored between 18 and 24, creating a gap between “experts” and “non-experts” on the continuum. After implementing these standards, a total of 101 participants from the Psychology 100 subject pool remained in the analysis: 18 experts and 83 non-experts.

As mentioned earlier, there was a large difference in the sample sizes of experts and non-experts in the current experiment, but this difference is ecologically valid. That is, experts should, in fact, be rarer than non-experts, as expertise involves having special knowledge that others do not possess. This ecological truism posed a problem for data collection, as only about 14% of participants tested were considered experts by the standards outlined above. Obtaining data from the varsity baseball team was an attempt to better balance the sizes the expert and non-experts groups for data analysis purposes. However, as will be discussed below, the data from the experts from the Psychology 100 subject pool could not be amalgamated with the baseball players’ data due to meaningful differences between the two samples.

Results. The first result that must be explained is the comparison between experts from the Psychology 100 subject pool and players from the Ohio State University varsity baseball team. The purpose of collecting data from the baseball players was to obtain more data that could be used in conjunction with the data provided by the participants considered to be experts from the Psychology 100 subject pool. In order to ensure that combining data from these groups was a valid step, I analyzed the baseball players’ ($n =$

24) scores on the 9-question abridged version of the 33-question expertise quiz, and compared these scores to experts' ($n = 18$) responses on the same 9 questions. A t-test revealed that the mean number of questions answered correctly by the experts ($M = 4.17$, $sd = 1.04$) and the baseball players ($M = 5.58$, $sd = 1.64$) differed significantly, $t(40) = -3.21$, $p < .01$, Cohen's $d = -1.03$. This difference in expertise quiz scores suggests that these two groups differed in their level of expertise, a distinction that proved to engender differences in the main dependent variable (hindsight bias). In light of this difference in expertise quiz scores, the following analysis of hindsight bias was carried out separately for experts from the Psychology 100 subject pool (referred to as "experts" from here on) and the baseball players. It should be noted, however, that the number of participants in the Psychology 100 expert hindsight group ($n = 8$) is too small to be able to confidently draw conclusions about the behavior of this group.

In order to assess the magnitude of hindsight bias demonstrated by the baseball players, experts, and non-experts, I found the proportion of baseball players, experts, and non-experts who answered each question correctly (foresight groups only). This measure represented the foresight values for baseball players, experts, and non-experts separately. Next, I subtracted the expertise-specific foresight proportion from the hindsight percentage each participant gave for each question. This deviation between foresight proportions and hindsight probabilities is the measure of hindsight bias. Finally, I computed a mean hindsight bias for the baseball players, experts, and non-experts, where a greater number indicates a greater deviation between foresight and hindsight values, and thus a greater hindsight bias. I compared these means using an analysis of variance (ANOVA) including all three expertise levels. Results of the ANOVA revealed that there

were significant differences between the three expertise groups, $F(2, 68) = 3.49, p < .05$, $\eta_p^2 = .09$. Tukey tests identified that the only significant difference was between the means of the non-experts ($M = 12.86$) and baseball players ($M = -2.59$), $p < .05$; the mean of the experts ($M = 10.51$) was not significantly different from that of the non-experts or the baseball players. Since the sample size of the expert group was rather small ($n = 8$), making it difficult to draw conclusions about this group, data from this group will not be discussed further. Recall that the purpose of collecting data from the baseball players was to increase the sample size of the expert group, since there were too few participants in this group. Since the baseball players' data could not be combined with that of the experts, the expert group remains too small to include in further analyses. Therefore, I decided to conduct a t-test between the means of the non-experts and the baseball players, eliminating the data from the experts. Results of the t-test revealed again that the non-experts and the baseball players displayed significantly different degrees of hindsight bias, $t(61) = 2.69, p < .01$, Cohen's $d = .75$.

The means of the non-experts' and baseball players' hindsight bias were not only significantly different from each other, but they also suggest that non-experts fell prey to the hindsight bias while the baseball players did not. In order to test whether the baseball players showed hindsight bias at all I conducted a one-sample t-test, comparing the mean of the baseball players' hindsight bias to zero. A hindsight bias value of zero would indicate that participants' hindsight probabilities did not deviate from the true proportion of foresight participants who answered each question correctly; in other words, participants did not think they "knew all along" answers that they truly would not have known in foresight. Results of the t-test revealed that baseball players' average hindsight

bias did not differ significantly from zero, $t(14) = -.44, p > .05$. The hindsight bias of the non-experts, however, was significantly different from zero, $t(47) = 4.87, p < .01$.

Discussion. Taken together, the results of Experiment 2 suggest that “super expertise” may shield individuals from the hindsight bias. Contrary to the results of Experiment 1, where hindsight bias increased with expertise, expertise reduced the hindsight bias in Experiment 2 to the point of non-existence. These seemingly contradictory results raise an interesting question about nature of the relationship between hindsight bias and expertise: does expertise exacerbate hindsight bias only up to a certain point, after which expertise is actually beneficial? While participants in Experiment 1 displayed greater hindsight bias as expertise increased, the comparison between experts and baseball players on the abridged version of the expertise quiz in Experiment 2 suggests that even the most expert Psychology 100 students may not have been experts in the strictest sense of the word. Shanteau (1992) noted that a true definition of what categorizes an “expert” is hard to find in the literature on expertise, but he argued that experts are typically recognized in their professions as possessing the skills necessary to perform domain-relevant tasks at the highest levels. He contrasted this functional definition of an “expert” with what is means to be “naïve,” namely possessing little or no skill in the domain in question. Finally, Shanteau (1992) called “novices” those who fall somewhere between naïve and expert decision makers. Novices (e.g., graduate students) have some advanced knowledge about the domain in question and often perform at “sub-expert” levels. Following these definitions, some researchers intending to examine so-called “experts” may have actually employed “novices” instead (e.g., Yates et al. 1991). According to Shanteau (1992) then, the “expert” participants in Experiment 1 more

accurately should have been considered novices rather than experts, whereas the baseball players in Experiment 2 certainly can be considered true experts. Again, this distinction between the “experts” and baseball players in Experiment 2 was supported by the baseball players’ significantly higher scores on the abridged expertise quiz.

Considering Shanteau’s (1992) definitions of naïve individuals, novices, and experts and the results from Experiments 1 and 2, more specific and nuanced conclusions can be drawn about the effects of expertise on hindsight bias. Participants in Experiment 1 demonstrated more hindsight bias as their score on the expertise quiz increased; as the classification of participants moved from “naïve” to “novice,” hindsight bias increased. However, as participants’ level of expertise increased even more, that is, as the classification of participants moved from “novice” to “expert” (in Experiment 2) hindsight bias dropped significantly. This pattern of results suggests an inverted U-shaped distribution of hindsight bias as expertise increases from a level of naiveté to a level of true expertise. Specifically, naïve individuals demonstrated a relatively small hindsight bias as compared to novices in Experiment 1. In Experiment 2, novices displayed a hindsight bias (they showed a hindsight bias that was significantly greater than zero), while experts did not demonstrate hindsight bias (the mean hindsight bias did not differ significantly from zero). Hindsight bias appears to be smallest for individuals who possess very little knowledge or a great deal of knowledge, while the bias is larger for those with intermediate levels of knowledge (i.e., novices).

Why should naïve individuals and experts be less-susceptible to hindsight bias in the domain in question? One possible explanation for the experts’ results could be that experts can identify the limits of their knowledge. Since experts possess more

information regarding the intricacies of their domain of expertise, they may be able to recognize more readily than novices that there is a great deal of information they do not know. For example, consider Professor M, a senior professor in the Quantitative Psychology department who attends weekly meetings in which members of the department present research. Professor M is recognized by members of his profession as being an expert in his particular sub-discipline. Throughout his tenure as a professor, he has listened to hundreds of presentations, some dealing with topics he knows a great deal about, others dealing with topics outside of his sub-discipline. Now imagine that Professor M were asked to take a 20-question quiz about Quantitative Psychology where he was given the questions and the correct answers to the questions and had to give the probability he would have known each answer had it not been provided (a typical hindsight task). Despite Professor M's expert-level knowledge in his Quantitative Psychology sub-discipline, Professor M has gleaned throughout the years that there is a great deal of information that he does not know within the domain of Quantitative Psychology. In fact, if Professor M knew everything there was to know about Quantitative Psychology, he probably would never attend another presentation. Therefore, when giving estimates of the probability that he would have known the answers to particular questions, perhaps Professor M would have more insight into what he knows and what he does not know. He may realize more readily than a novice (who is not aware of the many sub-topics within the domain) that there is so much information within his domain of expertise that he does not know about, and this insight may allow him to more readily indicate when he really would not have known something. While it

is simply a common aphorism (attributed to Aristotle), “the more you know, the more you know you don’t know” is supported by the results of Experiment 2.

While Aristotle would be pleased to see that the results of Experiment 2 support his conjecture, previous literature provides mixed evidence for the notion that experts have greater insight into the limits of their knowledge. Lichtenstein and Fischhoff (1977) showed that participants who answered the most general knowledge questions correctly in a multiple choice task demonstrated less overconfidence than did participants who performed worse on the task. However, the authors found no evidence that experts have a “quality of insight” (p.175) that leads them to assign confidence judgments more accurately. Specifically, multiple analyses of experts’ “resolution,” or the ability to discriminate when one’s answers are correct or incorrect, showed no difference between those scoring higher or lower on various multiple choice quizzes. Lichtenstein and Fischhoff’s (1977) conclusion was that “those who know more do *not* generally know more about how much they know” (p. 179, italics in original). In light of the distinctions delineated above between levels of expertise (Shanteau, 1992), however, the implications of Lichtenstein and Fischhoff’s findings for “true” experts are questionable. Participants in Lichtenstein and Fischhoff’s (1977) studies were undergraduates and graduate students in psychology, individuals who would be considered naïve or novices by Shanteau’s (1992) standards. Therefore, it is difficult to extrapolate from Lichtenstein and Fischhoff’s (1977) experiments whether experts, such as the baseball players in Experiment 2, have a “quality of insight” not exhibited by novices and naïve individuals.

It should be clear by this point that there are many hindrances to the study of the relationship between hindsight bias and expertise. The most obvious issue is the lack of a

clear definition of expertise. Some researchers studying hindsight bias or confidence in supposed experts and non-experts may not be studying “experts” at all. Instead, most of the experiments described herein demonstrate differences in biases as a function of relative expertise, rather than absolute expertise. Whether or not “absolute” expertise can be measured at all is an important question in and of itself; many attempts have been made to define expertise in a meaningful way, including identifying a “magic number” of hours spent practicing a task (e.g., Ericsson, Krampe, & Tesch-Romer, 1993; Gladwell, 2008) and delineating stages through which individuals must pass on their way to expertise (e.g., Benner, 1982; Dreyfus, 2004). The variability in the levels of expertise exhibited by participants in different experiments makes it difficult to compare results between studies and build a clear picture of how expertise influences the occurrence of various biases. In addition to this issue, or perhaps as a result of it, there is conflicting evidence in the literature regarding how expertise relates to hindsight bias. Some studies show that “experts” are susceptible to hindsight bias in their areas of expertise (e.g., Arkes et al., 1981; Guthrie et al., 2001), while others demonstrate that expertise protects individuals from falling prey to the hindsight bias (e.g., Gray et al., 2007; Guthrie et al., 2007). The combined results of Experiments 1 and 2 above suggest that an inverted U-shaped distribution best describes the relationship between expertise and hindsight bias, such that extremely high and extremely low levels of expertise lead to reduced hindsight bias, while intermediate knowledge exacerbates the bias. Additional evidence for an inverted U-shaped distribution is described below in Chapter 5, as is a discussion regarding absolute versus relative and objective versus subjective expertise.

CHAPTER 5

Experiment 3: Feeling-of-Knowing Explains the Relationship between Expertise and Hindsight Bias

Experiments 1 and 2 used objective measures of expertise (scores on a baseball quiz or holding a position on Ohio State's varsity baseball team) in order to explore the relationship between expertise and hindsight bias. Results thus far have suggested that expertise and hindsight bias are not linearly related, but rather that an inverted U-shaped distribution best describes this relationship; levels of hindsight bias are lowest for naïve individuals and experts, while those falling somewhere in between are most susceptible to the bias. Since true definitions of expertise continue to be disputed in the literature (e.g., Ericsson et al., 1993; Shanteau, 1992), it is difficult to identify which samples of participants in different studies can be considered experts and non-experts. Thus, most researchers who have conducted studies involving expertise have used relative levels of expertise, rather than absolute levels, when attempting to identify measures that vary with expertise. For example, Dawson et al. (1993) recorded physicians' confidence in their estimates of specific values associated with heart catheterization and found that overconfidence increased with expertise. Laypersons would likely consider all participants in the Dawson et al. (1993) study to be experts, while the researchers (some of whom were physicians themselves) made distinctions between levels of expertise

within the sample. Similarly, the baseball players who participated in Experiment 2 would most likely be considered by everyone to be experts in the game of baseball (assuming only skill increases, and not knowledge of the rules of the game, as players move from college ball to the minor and major leagues), while Psychology 100 participants who scored high on the expertise quiz may be considered baseball experts compared to their friends but not as compared to the Ohio State varsity baseball players. How can the same individuals be considered experts in some circles and non-experts in others? Anecdotally, this is a common occurrence. Mary may be the best cook her friends and family know, serving as the “resident expert” on everything culinary. It may not be until Mary’s first day of French pastry school that she realizes she is not an expert at all (a so-called “rude awakening”). When trying to study the effects of expertise on hindsight bias, is the important factor how expert Mary and her friends think she is (before her trip to the culinary school), or is objective expertise a more important factor? Is subjective expertise enough for participants to feel like they “knew it all along” when the questions are in subject areas they think they know more about? In other words, do participants simply have to *feel* like they know more about the subject in order to exhibit greater hindsight bias, or do they actually have to know more? These questions relate to the Jeopardy! scenario presented at the beginning of this paper. I feel like I know a lot about Spanish topics because I was a Spanish major in college. My husband feels like he knows a lot about literature because English is one of his college majors. While we each likely know more about these topics than do those who were not Spanish or English majors in college, and we know more about these topics than we do other topics, do we

actually know as much as we think we know in these areas, or do we feel like we know more than we actually do because information dealing with these areas is familiar to us?

In Experiment 3, I demonstrate that people think or feel like they know more about their areas of “expertise” than they actually do, and it is this discrepancy between what we know and what we think we know that leads to increased hindsight bias in one’s area of presumed expertise. Recall the results of Experiment 1 demonstrating that while baseball “experts” did not actually know much more than did non-experts without reading the essay about the 2006 MVP race (dogs group), experts nevertheless felt like they *would have* known this information without reading the essay to a greater extent than did non-experts. What accounts for this discrepancy between objective and subjective knowledge?

Glenberg and Epstein (1987) examined participants’ ability to assess their performance on an inference task after reading a text within or outside of their domain of expertise. Participants in this study were undergraduates enrolled in physics and music classes who had completed a minimum requirement of two university-level courses in either subject. Therefore, the term “expertise” is used in this study to indicate some experience within a particular domain (many participants also participated in other domain-related activities such as music lessons or working as a laboratory assistant in a physics lab). Participants were asked to read 16 texts, 8 having to do with music theory and 8 dealing with physics. After reading the texts, the participants had to answer a series of “probes” asking them, among other things, to a) rate their confidence in their ability to draw correct inferences from each text, b) actually answer the inference questions, and c) rate their confidence that they had answered each particular inference

question correctly. One of Glenberg and Epstein's (1987) main purposes was to test the "expertise hypothesis," that experts would be better calibrated in their area of expertise. However, the authors posited that perhaps participants would only be able to predict that they would perform better in their area of expertise than they would in the other subject area, and not how well they would perform *within* each domain. In this case, participants would be well-calibrated overall (they would assign lower confidence ratings for their ability to make correct inferences outside of their area of expertise and higher ratings for questions in their area of expertise), but they would be poorly calibrated *within* each domain.

There were a number of important results from this experiment. First, Glenberg and Epstein (1987) found, as expected, that the more physics or music courses participants had completed, the more confident they were that they would be able to draw correct inferences from the respective texts. Furthermore, the increase in confidence imparted by the additional classes completed was greater within domains than between domains. That is, taking more physics classes increased participants' confidence that they could draw correct inferences from physics texts to a greater extent than it increased their confidence that they could answer the music questions (and vice versa for the music students). Participants who had taken more courses in a particular subject also answered more questions correctly in their respective areas of expertise (i.e., number of courses taken was positively related to performance). Taken together, these results suggest that, across domains, participants were well-calibrated; they appropriately assigned higher confidence ratings to texts within their domain of expertise than to texts outside of their area of expertise. However, participants were not well-calibrated within domains. More

interestingly, participants' level of knowledge within their domain was negatively associated with (for physics students) or unrelated to (for music students) calibration in that domain. These results suggest that "experts" are aware that they know more in their area of expertise than they do outside of it, but they have less insight into how much they know within their area of expertise.

Glenberg and Epstein (1987) argued for a "self-classification hypothesis" to explain their results. Specifically, the authors suggested that participants self-classified themselves as more expert in one subject or the other (physics or music), and that the students therefore believed that they would do better on the inference task in their area of self-classified expertise. The students were correct in these classifications and predictions, since they did better on the inference tasks in their area of self-classified expertise and predicted that they would. However, Glenberg and Epstein (1987) posited that confidence is based on these general self-classifications of expertise and not on the specific context of the texts for which the inferences would be made. Glenberg and Epstein (1987) tested the self-classification hypothesis by simulating predictions of participants' confidence ratings for the inference task in each domain based on their experience with the domain (number of physics classes and number of music classes) and comparing these values to participants' reported confidence for the inference task in each domain. Results showed that the mean simulated confidence rating was not significantly different from the mean confidence rating provided by participants. This finding has major implications; participants based their confidence judgments almost exclusively on their self-classifications of expertise with a particular domain, without taking into account the specific content of the text in question. That is, participants did not factor into their

confidence ratings how much they knew about the unique content of each text, but rather based their confidence ratings on how much they knew about the domain from which the text was drawn. Glenberg and Epstein (1987) concluded that individuals' assessment of their knowledge in their domain of expertise is not "fine-grained" enough to be able to predict differences in performance within that domain.

Glenberg and Epstein (1987) suggested that familiarity with a topic may also be responsible for the poor calibration demonstrated by self-classified experts. If participants base their confidence ratings on familiarity with a topic in their domain of expertise and not the specific information provided in a particular text, participants will be poorly calibrated. In fact, Glenberg and Epstein (1987) showed that familiarity ratings were highly correlated with confidence ratings, which were essentially substitutable for participants' hypothesized self-classifications of expertise. Glenberg, Sanock, Epstein, and Morris (1987) tested a "domain familiarity hypothesis," by which domain familiarity would predict confidence ratings. The authors asked three different groups of participants (all expected to have similar domain knowledge) to read 15 texts and subsequently a) give familiarity and confidence ratings, b) recall information from each text, and c) answer inference questions pertaining to the texts. Consistent with Glenberg and Epstein's (1987) findings, familiarity and confidence were highly correlated, but the correlation between familiarity and performance on the inference task was extremely low. In an additional experiment testing the domain familiarity hypothesis, Glenberg et al. (1987) found that asking participants for familiarity ratings by presenting either a verbatim restatement of the text's central argument or a paraphrase of this argument (verbatim-paraphrase manipulation) affected familiarity ratings. Specifically, the authors

found that participants rated verbatim statements as more familiar than paraphrases of the arguments in the text. Glenberg et al. (1987) capitalized on this finding in yet another experiment; when familiarity was manipulated via the verbatim-paraphrase manipulation, participants reported differences in familiarity (i.e., verbatim statements were again rated as more familiar than paraphrases) but no differences existed between confidence ratings or performance on the inference task between the groups. The authors concluded that although evidence for the domain familiarity hypothesis was mixed when the verbatim-paraphrase manipulation was used, familiarity with a text nevertheless correlated with confidence in their earlier study (explained above) and in Glenberg and Epstein's (1987) study. Additionally, Glenberg et al. (1987) argued that since familiarity with a specific statement (which was what was manipulated by the verbatim-paraphrase manipulation) did not affect confidence, this actually does support the notion that confidence is not related to information contained in specific questions, but rather it is related to domain familiarity.

Glenberg and Epstein (1987) and Glenberg et al. (1987) provided evidence that individuals may use their general familiarity with a particular domain as a cue for their ability to answer specific questions within that domain. However, domain familiarity clearly is not a particularly valid cue, as evidenced by participants' poor calibration within their domain of expertise. In addition to evidence pointing toward individuals' use of a domain familiarity strategy in predictions of future performance, there is also evidence that individuals use this strategy in their assessments of their prior knowledge, that is, how much they knew prior to receiving outcome information. As mentioned in Chapter 3, Werth and Strack (2003) and Marks and Arkes (2009) both explored the

effects of familiarity on the hindsight bias or “knew-it-all-along” effect. Werth and Strack (2003) suggested, like Glenberg and Epstein (1987) and Glenberg et al. (1987), that familiarity and confidence are related. Werth and Strack (2003) argued that experienced confidence or certainty when making inferences about an answer leads to feelings of familiarity and, consequently, a greater feeling that an individual “knew it all along.” In two experiments the authors demonstrated that participants who experienced greater feelings of confidence or perceptual fluency assimilated more to estimates provided as answers to questions. While not specifically tested by Werth and Strack (2003) it seems likely that familiarity also increases processing fluency. If individuals are more familiar with a particular domain, the terminology, people, and events dealing with this domain are likely processed more fluently, leading to increased confidence that the information is something that an individual knew already. Marks and Arkes (2009) directly manipulated familiarity by varying the content of the material presented to participants and demonstrated that participants only displayed hindsight bias when the information was familiar; obscure information did not engender the bias. Source confusion was implicated in Marks and Arkes (2009) studies as being responsible for the effect of familiarity on hindsight bias. Specifically, the authors argued that participants could not identify the source of their current knowledge when the information was familiar, thus making it difficult for participants to discount the specific information contained in the essay when asked to do so. This explanation fits well with Glenberg et al.’s (1987) domain familiarity hypothesis; participants judged whether information was new or old based on their familiarity with the domain of the information, not on the specific information presented. When participants in Marks and Arkes’ (2009) study read

an essay coming from a familiar domain, they felt familiar with most of the information, and thus they answered about as many questions correctly on a quiz as participants who were trying to do their best on the quiz. Rather than identifying correctly whether they already knew the specific piece of information presented, participants in Marks and Arkes's (2009) study likely felt a sense of familiarity with most of the information, thereby inferring that they actually knew that specific information all along.

As mentioned earlier, Glenberg and Epstein (1987) argued for a "self-classification hypothesis," which states that individuals classify themselves as relatively more expert in some subjects than in others and proceed to use this self-classification when predicting how they will perform on specific questions in their area of presumed expertise. While Glenberg and Epstein (1987) showed that self-classifications of expertise were somewhat justified (i.e., participants answered more questions correctly in their area of expertise than in other areas), it is clear that the amount of information individuals feel they know *within* their area of self-classified expertise is overestimated (as argued in the domain familiarity hypothesis). This discrepancy between what people think they know and what they actually know is the crux of the hindsight bias. If individuals are more familiar with information in their domain of self-classified expertise, and if this feeling of familiarity leads them to feel like they know more than they actually do in the domain, then when they are asked to determine whether or not they would have known something had they not been given the answer, they are going to give higher probability estimates in their domain of self-classified expertise than they would in other domains. Furthermore, these higher probability estimates will be more discrepant from individuals' actual levels of knowledge if specific knowledge does not equal domain

familiarity. Recall the example of Mary, the resident food expert. Mary likely has heard of cream of tartar, “proofing” bread, and “blind baking,” but that does not mean that she could correctly identify what each of these ingredients or techniques does. If Mary is asked how likely it is that she would have known that cream of tartar is an ingredient used to stabilize egg whites, she would assign a high probability to this question if her probability estimate is based on familiarity. However, if Mary were actually asked what ingredient is often used to stabilize egg whites, she may not actually know that the answer is cream of tartar. At the beginning of this chapter, the question was posed as to whether subjective expertise is enough to exacerbate the hindsight bias. It would seem that if subjective expertise is related to familiarity, and if familiarity is related to hindsight bias, then subjective expertise should be enough to exacerbate the hindsight bias.

Literature on “feeling of knowing” (FOK) may help to further elucidate whether subjective expertise is enough to produce greater hindsight bias. A number of researchers (e.g. Costermans, Lories, & Ansay, 1992; Koriat, 1993; Koriat & Lieblich, 1977; Metcalfe, 1986; Nelson & Narens, 1990; Reder, 1987; 1988) have studied participants’ ability to provide FOK judgments after failing to recall a specific target. FOK judgments typically require participants to express on a particular scale (e.g. 1-9 as in Nelson & Narens, 1980; or 0-100 as in Koriat, 1993) the chance that they would be able to identify a target or correct answer, and this FOK judgment is typically elicited only for questions that participants failed to answer correctly previously (but, see Koriat, 1993, for a discussion of problems with this technique). While it may seem obvious that the extent to which participants “feel” like they know the answer to a particular question is

intimately tied to hindsight bias, the FOK literature and the hindsight bias literature rarely mention the other line of experimentation (see Hoch & Lowenstein, 1989, for an exception). In Experiment 3 I link the findings in these two literatures by demonstrating that greater hindsight bias results from greater expertise because feeling of knowing is greatest in one's area of presumed expertise. Specifically, I argue that it is not what one *actually* knows in one's area of presumed expertise that contributes to increased hindsight bias, but rather it is what one *feels like* he knows in a particular domain that accounts for the discrepancy between foresight accuracy and hindsight judgments. Since expertise in Experiment 3 is based on participants' rankings of their areas of expertise, Glenberg and Epstein's (1987) self-classification hypothesis comes into play. Additionally, the domain familiarity hypothesis (Glenberg et al., 1987) described above is expected to contribute to the predicted relationship between feeling of knowing and hindsight bias.

There are differing schools of thought within the FOK literature that identify different determinants of FOK judgments. The distinction between the possible determinants of FOK judgments is important for understanding the predictions of Experiment 3. Those supporting the *cue-familiarity* account of FOK (Metcalf, Schwartz, & Joaquim, 1993; Reder, 1987) argue that it is the familiarity of the "pointer," or the memory cue, that drives FOK judgments. In studies addressing the cue-familiarity hypothesis, participants typically give quick or "preliminary" (Reder, 1987) FOK judgments on the basis of the pointer. The *cue-accessibility* account of FOK (Koriat, 1993, 1995) posits that feeling-of-knowing judgments are based on the partial retrieval of cues that are relevant to the question at hand. When participants fail to correctly answer a question and subsequently are asked to give FOK judgments, these judgments are

assumed to depend on the amount and strength of partial cues encountered during the previous attempt to answer the question. In experiments testing the cue-accessibility hypothesis, participants typically attempt to answer a particular question and then give FOK judgments indicating the chance that they would be able to recognize the correct answer in a subsequent task. FOK ratings are usually untimed in these experiments.

Korait and Levy-Sadot (2001) attempted to reconcile these differing explanations of the process underlying FOK judgments. Through a series of experiments, Koriart and Levy-Sadot (2001) demonstrated that cue-familiarity and cue-accessibility accounts of FOK are not in conflict, but rather they function in a “cascaded” (p. 43) manner to influence FOK judgments. Specifically, cue familiarity exerts its influence in the preliminary stages of memory retrieval, while cue accessibility impacts FOK judgments in later stages of retrieval. Operationally, FOK judgments are presumed to be based on cue familiarity if participants are not given much time in which to make FOK judgments; FOK judgments that incorporate metacognitive information from the retrieval of partial cues during memory search can only be attained if participants are given enough time to actually search their memories for relevant information. In Experiment 3 FOK judgments were made rapidly; participants were given only 2 seconds to make each FOK judgment. This time constraint was implemented to ensure that participants made FOK judgments based on cue familiarity and not cue accessibility. Since the current hypothesis is that familiarity in a particular domain is what leads to increased hindsight bias, tapping only into domain familiarity is an important feature of the experimental procedure. In the 2 seconds participants had to give each FOK judgment, it is unlikely that participants would have been able to focus on more than memory “pointers” in the

questions themselves to evaluate their feeling of knowing for each question. It follows, then, that participants likely based their FOK judgments on their familiarity with the terminology, people, and places in the questions. The more familiar participants were with the information contained in the question, the higher their FOK ratings have should been. Therefore, FOK ratings should vary with ranked expertise. Furthermore, if familiarity with a domain leads individuals to feel like they know more in that domain, then probability estimates of the likelihood that participants would have know the answer to a question had they not been given the answer should be higher in more familiar domains. Following the domain familiarity hypotheses (Glenberg & Epstein, 1987), increased familiarity with a domain, and consequently increased feeling of knowing, should not necessarily increase accuracy on specific questions in the domain. Together, all of these factors lead to the hypothesis that average FOK should increase with self-classified (ranked) expertise in a domain, while average accuracy on questions within a domain will not follow in kind. An explicit exposition of each hypothesis for Experiment 3 is presented after the Procedure section below, as the hypotheses can be more clearly explained in light of the specific procedures used in the experiment.

Method

Participants. One-hundred fifty-three undergraduates enrolled in Psychology 100 at The Ohio State University participated in this study. Each student received course credit for his or her participation.

Materials. Participants heard a series of 100 questions dealing with history, literature, science, mathematics, and geography (see Appendix D for questions). There were 20 questions from each of the five domains. The questions were recorded onto a

cassette tape and played on a tape recorder to ensure consistent spacing between questions and to minimize experimenter influence. After hearing each question, participants typed FOK ratings onto a computer using the MediaLab program.

In addition to the 100 questions presented on the audio cassette, participants saw the same questions in the same order on the computer using the MediaLab program. According to the experimental condition to which they were randomly assigned, participants were asked to type answers to the questions or to type hindsight probabilities in the space provided.

Procedure. First, participants ranked the five domains (history, literature, science, math, and geography) in order from the domain they thought they knew the most about (most expert) to the domain they thought they knew the least about (least expert) by dragging and dropping the domains from the left side to the right side of their computer screens. Prior to subjects hearing the experimental questions, the experimenter told the participants how to give FOK ratings, explaining that they are quick, immediate reactions to the questions themselves, without necessarily thinking about what the answers to the questions are. Participants were told that they were to give FOK ratings on a scale of 0 to 9, where 0 meant that, upon hearing the question, they felt like there was no way they would know the answer to the question, and 9 meant that, upon hearing the question, they felt like they definitely would know the answer to the question. Participants were given two seconds after the question was completed to give each FOK rating before the next question began. Participants listened to four practice questions designed to get them accustomed to the pace at which the questions would be read, ensure that the tape was at an appropriate volume, and confirm that they fully understood

how to give FOK ratings. After hearing the four practice questions and giving practice FOK ratings, the actual experiment began. For approximately 15 minutes, participants listened to the 100 experimental questions and gave immediate FOK ratings without pause. After the FOK portion of the experiment concluded, participants moved on to the foresight or hindsight portion of the experiment. Participants were randomly assigned by session to one of the experimental groups so that the experimenter could give condition-specific instructions to the participants. Participants in the foresight group saw the same 100 questions that they heard on audio tape in the same order and were asked to type the answer to the question in the space provided. They were told that if they could not answer the question, they should try to make an educated guess or type “I don’t know” if they could not provide an educated guess. Participants in the hindsight group saw each question that they heard on the audio tape in the same order, and they also saw the answer to each question in parentheses after the question. For example, participants in the hindsight group saw the following question:

“What is the capital of Australia? (Canberra)”

Participants in the hindsight group were instructed to rate on a scale of 0-100 the probability that they would have answered the question correctly had we not provided them with the correct answer.

Hypotheses. I predicted three main findings that, taken together, would allow me to argue that subjective expertise leads to increased hindsight bias because subjective expertise leads to greater FOK values, but not necessarily greater accuracy.

H1: *Average FOK ratings will increase with ranked expertise.* I predicted that the FOK ratings participants gave for the 20 questions in each domain would correspond to

their ranking of the domain, such that the domain ranked highest would have the highest average FOK rating, and the domain ranked lowest would have the lowest average FOK rating. This expected result would demonstrate that participants *felt* like they knew more in the domains in which they thought they were most expert. Koriat and Levy-Sadot (2001) demonstrated a similar finding in their Experiment 1, however they showed that FOK judgments were higher when familiarity with a memory pointer was higher. In the current experiment, I used ranked expertise as a measure of familiarity. Presumably, participants should be more familiar with information in the domains they ranked higher in expertise, thus I expected findings similar to those of Koriat and Levy-Sadot even though the independent variable was slightly different.

H2: *Subjective expertise will not indicate objective expertise.* I predicted that subjective expertise would afford only minor increases in accuracy to those who ranked a particular domain most expert as compared to those who did not rank the same domain as most expert. While participants likely know somewhat more in the domains they ranked highest as compared to participants who did not rank those domains highest (Glenberg & Epstein, 1987), accuracy was predicted be low in general. Unlike Bradley's (1981) measure of poor accuracy for experts, where experts failed to score above chance on a true/false quiz, there is no statistical test that allows me to indicate objectively whether "experts" perform well or poorly on a fill-in-the-blank quiz requiring correct recall. Nevertheless, I predicted that subjective expertise would not result in high accuracy.

This prediction is supported by the fact that in Experiment 1, increased baseball expertise did not predict a significant increase in scores on a true/false quiz about baseball current events when participants were not provided with information about the

events (dogs group). This discrepancy in the performance of baseball experts on the different baseball quizzes may indicate that expertise is not as generalizable as one might expect. That is, superior performance in one subset of questions in a particular domain does not necessarily mean that superior performance should be expected on a different subset of questions in the same domain. Since participants in Experiment 3 were asked to rank their areas of expertise based solely on the name of the domain (i.e. “math,” not “math definitions” or “mathematical computation,” and “science,” not “biology” or “astronomy”), participants may have been expected to overgeneralize their expertise in a particular domain during the ranking process. Such overgeneralization of one’s knowledge would be in line with Glenberg et al.’s (1987) domain familiarity hypothesis; when individuals are familiar with a particular domain, they predict better performance on specific questions within that domain than is warranted. However, overgeneralization during the domain ranking task would not excuse participants from giving FOK ratings on specific questions that are too high for one’s actual knowledge. In Glenberg et al.’s (1987) tasks, participants made predictions about their future performance on specific questions, presumably based on familiarity with the domain. In those tasks, participants did not know what the specific questions were when making predictions of their future performance. However, in the current experiment participants knew the specific question for which they were giving FOK judgments, so they should have corrected for overestimations due to domain familiarity on specific questions. That is, even if participants ranked “science” highest despite the fact that they really were experts only in astronomy, they should have corrected for this overgeneralization of expertise in the ranking task by giving lower FOK ratings on science questions not dealing with

astronomy. Such a correction would lead participants to lower their FOK ratings in response to the specific questions they were asked. If participants engaged in this type of correction, I would not expect to see the relationship between ranking and FOK predicted in Hypothesis 1.

One might also argue that if participants do not score high on every test in their area of expertise, then they are not actually experts in this domain. While this is a valid criticism of my characterization of expertise in the current series of experiments (in fact, criticisms of most categorizations of expertise are warranted, since expertise is generally not well-defined), such a criticism does not explain the finding that expertise exacerbates hindsight bias (up to a certain point). The overgeneralization of expertise described above is important for the proposed mechanism underlying hindsight bias in those who are more expert in a particular domain: as expertise increases, individuals *feel* like they know more than they actually do, and this *unwarranted* feeling of knowing results in a greater discrepancy between what one knows in foresight and what he thinks he would have known in hindsight. Participants in Experiment 1 who scored higher on the 33-question quiz about baseball rules clearly knew a good deal about baseball. However, this knowledge about baseball rules does not necessarily translate to knowledge about current events in baseball. Importantly, however, those who know a great deal about baseball rules, and who would be expected to place “baseball” high on a list of domains ranked by expertise, likely feel like they know a good deal about baseball current events as well. Baseball players, baseball teams, baseball scores, baseball stats, and other subtopics of baseball information are probably highly familiar to those scoring high on the 33-question baseball rules quiz, leading these “experts” to misjudge how much they actually

would have known about the baseball current events delineated in the 2006 MVP race essay before reading the essay. A similar effect should emerge in Experiment 3: ranked expertise may not translate to a substantial increase in accuracy.

H3: *Subjective expertise will result in increased hindsight bias.* Finally, I predicted that the degree of hindsight bias participants demonstrated would increase with ranked expertise. The discrepancy between what participants actually knew (H2) and what they felt like they knew (H1) was predicted to increase as self-classified expertise (rank) increased.

Results. I made three predictions regarding the results of Experiment 3. First, I hypothesized that average FOK ratings would increase with ranked expertise. In order to test this hypothesis, I conducted a repeated measures analysis of variance (ANOVA) with domain rank as the independent variable and average FOK as the dependent variable. For each participant I calculated the average FOK value for the domain at each rank. For example, suppose participant X ranked the five topics in the following order (from most expert to least expert): science; math; history; geography; literature. Participant X's average FOK value for science (i.e., the average of all 20 FOK values participant X gave for each science question) would be calculated, and, since participant X ranked science highest, the average FOK value for science would be assigned as participant X's FOK value for rank = 1. Similarly, participant X's average FOK value for literature would be calculated, and, since participant X ranked literature lowest, the average FOK value for literature would be assigned as participant X's FOK value for rank = 5. This same process was repeated for each participant for each domain, resulting in each participant having an average FOK rating for the domain at each of the five rank levels.

Participants' average FOK values at each rank level were then subjected to a repeated measures ANOVA. The ANOVA revealed that the test of within-subjects effects was significant, $F(4, 608) = 89.73, p < .01, \eta_p^2 = .37$, demonstrating that as rank increased average feeling of knowing increased as well (Figure 4). As predicted, average FOK increased with participants' ranking of the different domains from least to most expert.

The second hypothesis for Experiment 3 was that subjective expertise would not necessarily indicate objective expertise. I predicted that although subjective expertise would likely afford participants some additional knowledge in their domains of presumed expertise, participants nevertheless would not be very accurate when answering the questions in their most expert domains. As a measure of participants' accuracy, I considered the 20 questions from each domain as separate quizzes and gave participants a "score" on each quiz. That is, each participant answered a certain number of questions correctly out of the 20 questions from each domain, and this number of questions answered correctly served as a participant's score for each domain (science, history, literature, geography, and history). Participants' scores were then categorized by rank in the same way participants' average FOK ratings were categorized by rank in the test of Hypothesis 1. Consider again participant X, who ranked the domains in the following order (from most expert to least expert): science; math; history; geography; literature. Participant X's score on the science "quiz" (i.e., the number of the 20 science questions participant X answered correctly) would be the value associated with rank = 1. Similarly, participant X's literature score would be the value associated with rank = 5. This same process was repeated for each participant for each domain, resulting in each participant having a quiz score for the domain at each of the five rank levels. Participants' quiz

scores at each rank level were then subjected to a repeated measures ANOVA. The ANOVA revealed that the test of within-subjects effects was significant, $F(4, 608) = 22.82, p < .01, \eta_p^2 = .13$, demonstrating that as rank increased quiz scores increased as well (Figure 5). While the results demonstrate that, as predicted, subjective expertise afforded participants with increased accuracy, it is important to note that participants scored very low on the quizzes overall. As can be seen in Figure 5, participants' scores on the individual topic quizzes ranged from 2.17 (rank = 5) to 3.86 (rank = 1); participants answered less than four questions correctly out of 20 in the domain they ranked as most expert! This is quite poor performance by any standard. The implications of this poor performance will be discussed later.

Finally, I predicted in Hypotheses 3 that subjective expertise would result in increased hindsight bias. That is, participants should demonstrate a greater discrepancy between what they know and what they think they would have known as the domains increase in rank from least to most expert. To calculate the magnitude of hindsight bias displayed by each participant, I matched hindsight participants to foresight participants on the basis of the domain they ranked at each level. Imagine participant X was in the hindsight group and ranked the domains, as before, in the following order (from most expert to least expert): science; math; history; geography; literature. In order to calculate participant X's hindsight bias value, I first would have to select only participants who, like participant X, ranked science as most expert. Once I restricted the sample of foresight participants to those who ranked science highest, I would calculate the proportion of this subset of participants who answered each science question correctly. Using this process I would obtain the proportion of foresight participants who ranked

science highest that answered each of the 20 science questions correctly. From here participant X's hindsight bias value would be computed by subtracting participant X's hindsight probability (a number between 0 and 100 that indicated how likely it was that participant X would have known the answer to the particular question had I not provided the correct answer) on a particular science question from the proportion of foresight participants who answered that same question correctly. The average of these 20 difference scores would be participant X's hindsight bias value for rank = 1. Similarly, participant X's hindsight bias value for rank = 5 would be computed by a) selecting the subset of foresight participants who, like participant X, ranked literature lowest in expertise, b) finding the proportion of this subset of participants who answered each literature question correctly, and c) subtracting participant X's score on each literature question from the respective foresight proportion for each literature question. This process was carried out for each domain at each rank level for each participant, resulting in each participant having a hindsight bias value for each of the five rank levels. The method of partitioning foresight participants into subgroups based on the topic they ranked at each level ensures that the degree of hindsight bias obtained takes into account the differences in knowledge participants had regarding the domain at each rank level. In other words, I wanted to make sure that the only participants entering into the calculation of the hindsight bias were those who claimed to have similar levels of knowledge in a particular topic.

After obtaining a hindsight bias value for each participant at each rank level, the hindsight bias values were subjected to a repeated measures ANOVA. The ANOVA revealed that the test of within-subjects effects was significant, $F(4, 304) = 23.62, p <$

.01, $\eta_p^2 = .24$, demonstrating that as rank increased hindsight bias increased as well (Figure 6). As predicted, hindsight bias was exacerbated by subjective expertise; as rank increased from least to most expert, the difference between what participants actually knew and what they thought they would have known increased.

Discussion. Like Experiments 1 and 2, Experiment 3 also explored the relationship between expertise and hindsight bias. However, Experiment 3 had some important features that differed from the previous two experiments. First, Experiment 3 was a within-subjects test of the expertise-hindsight relationship. While the previous studies looked at differences between groups, the materials and procedure employed in Experiment 3 allowed for a demonstration of the effects of expertise on hindsight bias within the same person. The results of Experiment 3, therefore, are a clear demonstration that hindsight bias is affected by expertise even within a single individual. Second, Experiment 3 used a subjective measure of expertise, namely the ranking of the domains, rather than an objective measure of expertise (the 33-question expertise quiz). The finding that hindsight bias differed with participants' ranking of their own expertise supports the notion that individuals simply have to *feel* like they know a good deal of information in a domain in order for hindsight bias to be exacerbated. Indeed, the subjectivity of participants' expertise was confirmed by their abysmal scores on the individual sub-quizzes. Finally, the important inclusion of feeling-of-knowing ratings in Experiment 3 provided important information not present in Experiments 1 and 2 about the psychological processes involved in the occurrence of hindsight bias.

The results of Experiment 3 raise a number of important issues about how individuals classify their own levels of expertise, how this classification relates to actual

performance, and how both of these factors relate to hindsight bias. First, how do people categorize themselves as experts in some areas and non-experts in others? In Experiment 3, participants were asked to rank different domains in the order they thought best reflected their levels of expertise in these domains. It is conceivable that some participants considered themselves to be experts in all of the domains or completely naïve in all of the domains. Or, some participants could have considered themselves to be equally as expert in two domains and completely naïve in the other three domains. Any such relationship between levels of expertise in the five domains could exist for each participant. Nevertheless, participants were forced to rank the domains, leaving no possibility for ties between the domains and no way for participants to express expertise or naiveté in all of the domains. While the nature of the ranking task could be seen as a limitation of Experiment 3, it is informative in its own right. The ranking task essentially asked participants to consider their *relative* expertise between the domains rather than their *absolute* expertise in each domain. The implication of thinking about the domains in a relative rather than an absolute sense is that participants had to make no claims as to how much they knew in each domain compared to other people. Some complicated issues arise when individuals compare themselves to other people, most notably, the majority of people think they are better than average (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995). If participants were asked to rate their expertise in each domain on a Likert-type scale, they may have engaged in between-person comparisons, concluding that they were better than average in all domains and giving artificially high ratings to reflect this conclusion. This data would be far less informative because it would be based on a combination of factors, only one of which may be expertise.

Another benefit to assessing participants' relative expertise by asking them to rank the domains is that this procedure is probably less likely to induce self-presentational concerns. When participants were asked to rank domains in order from most to least expert, they were not actually declaring that they were experts in any of the domains. As mentioned above, the ranking procedure allows for participants to be relatively naïve in all the topics while still satisfactorily completing the ranking task. Therefore, participants were less likely to feel like they had to perform at a certain level than they would if they were to give a Likert-type rating for each domain. That is, ratings may induce self-presentational concerns to a greater extent than rankings. This is an important point because previous research has suggested that self-presentation is related to hindsight bias (e.g., Campbell & Tesser, 1983; Musch & Wagner, 2007), such that greater self-presentational concerns lead to greater hindsight bias. It should be noted at this point that placing the ranking task before the collection of the other dependent variables could be seen as a limitation of Experiment 3. One might think that asking participants to rank domains in the order of their expertise would prime them to answer the subsequent questions in a manner that was consistent with their ranking. While this is a valid concern, Bradley (1981) also asked participants to rank domains in order of their expertise either before or after giving confidence ratings for questions in each of the domains. The placement of the ranking task in the experiment made no impact on participants' confidence ratings, even though Bradley (1981) grouped the questions in the confidence-rating section by domain and placed a heading at the top of each set of questions identifying the domain from which they came. Bradley's finding suggests that placing the ranking task at the beginning of Experiment 3 did not lead participants to give

FOK ratings or hindsight probability estimates that were in line with the ranking order. Furthermore, participants' rankings varied significantly with the number of questions they answered correctly at each rank level; it seems unlikely that ranking could induce participants to answer more or fewer questions correctly depending on where they ranked each domain. There would be no reason to expect that simply performing the ranking task would affect some variables and not others.

Given that ranking domains only provides information about participants' relative subjective expertise in each domain, there was no objective classification of participants' levels of expertise as there was in the previous two experiments. Nevertheless, participants in the foresight groups answered all 100 questions, allowing for an analysis of accuracy for the domain participants ranked at each level. While the results showed that subjective relative expertise was related to accuracy, performance was extremely poor overall (participants answered less than 4 questions correctly in their highest ranked domain). This is an interesting finding, since such poor performance overall nevertheless resulted in significant differences in quiz scores as a function of rank. This suggests that participants had enormous insight into their relative expertise between domains. Glenberg and Epstein (1987) found a similar result in participants' confidence ratings; while participants were poorly calibrated within a domain, they were well-calibrated *across* domains. That is, they were able to accurately account for differences in expert versus non-expert domains. What is impressive about participants in Experiment 3 is how finely-tuned their insight actually was as to their relative expertise between domains. With a difference of fewer than 2 questions separating participants' highest and lowest ranked domains, there was a significant relationship between sub-quiz scores and ranked

expertise. This finding provides strong evidence for the validity of participants' own rankings of subjective relative expertise.

While participants in Experiment 3 knew very little overall, they certainly felt like they knew a lot. Feeling-of-knowing ratings, like accuracy, were related to ranked expertise. It is difficult to conclude solely from the feeling-of-knowing ratings and sub-quiz scores that participants felt like they knew *more* than they actually did, since the two variables were not in the same metric. However there is evidence to suggest why there would be a discrepancy between what participants actually knew and what they *felt* like they knew. The cue-familiarity hypothesis (Metcalf et al., 1993; Reder, 1987) states that FOK ratings are based on individuals' familiarity with memory "pointers" or cues in the questions themselves. FOK ratings are likely to be based on cue familiarity if participants are given only enough time to focus on aspects of the questions rather than on information retrieved from memory that is relevant to the answers. Since participants were allowed only 2 seconds to give an FOK judgment for each question, these judgments should have been products of participants' familiarity with terminology in the question. For example, when a participant heard the question "What is the name of the main character in the J.D. Salinger novel 'The Catcher in the Rye'?", the only processing the participant could have done in 2 seconds was to quickly assess whether or not he was familiar with "The Catcher in the Rye." If he was familiar with this novel, he should have given a relatively high FOK rating, since he would have felt that, given more time, he would be able to recall the answer. If the participant was unfamiliar with this novel, he should have given a low FOK rating, because he would have felt that, even given the time, he would not be able to recall the answer. Since both FOK ratings and accuracy

were related to ranked expertise, a participant's average FOK rating for each rank level is likely related to his accuracy at that rank level. However, this relationship certainly is not perfect. A participant could think indefinitely and never recall the correct answer even if he is familiar with "The Catcher in the Rye." It is clear from participants' poor overall accuracy that, for most questions, the participants did not eventually recall the answers. Poor accuracy accompanied by high FOK ratings is consistent with Glenberg et al.'s (1987) domain familiarity hypothesis. If participants' FOK ratings were based on familiarity with a particular domain (as argued in the cue-familiarity hypothesis), but accuracy was based on specific knowledge within a domain, a disconnect similar to the one Glenberg et al. (1987) found between within-domain confidence and accuracy should be expected for within-domain feeling of knowing and accuracy.

Recall that when participants were asked to rank the five domains in order of expertise, they were simply given one-word identifiers for each domain (math, literature, history, geography, and science). Of course there are many sub-topics within each of these domains with which participants may be more or less familiar. Nevertheless, participants were forced to rank these domains according to their expertise in the absence of information regarding what specific topics within the domains they would be asked about. Necessarily, participants' rankings were based on extremely general considerations of familiarity with each of the domains. As discussed above, feeling-of-knowing ratings were also likely based on participants' domain familiarity, but these ratings should have been more informed than the rankings were. That is, participants knew the specific questions for which they were giving FOK ratings, but the time limit imposed on participants' FOK judgments allowed them to make only superficial

judgments about their ability to answer each question. Unlike rankings and FOK judgments, hindsight probability estimates were given under full disclosure and without any time constraints. Therefore, even though participants were forced to give rankings based on generalizations of their expertise with each domain, they should have corrected for any overgeneralizations they made during the ranking procedure when they were asked to give hindsight probability estimates for specific questions. If participants appropriately engaged in such a correction procedure, hindsight bias should have been unrelated to ranked expertise. That is, regardless of how participants' ranked each domain, they should have made hindsight probability estimates based on each specific question. For example, even if a participant ranked science highest because he knows about certain topics in science more than he knows about certain topics in the other domains, when answering specific science questions, the participant should be able to accurately judge whether he would have known that specific science question; this specific hindsight estimate should be independent of the fact that the participant ranked science highest. However, the results of Experiment 3 showed that hindsight bias bore a significant relationship with ranked expertise; as ranked expertise increased, so did hindsight bias.

The results of Experiment 3 showed that sub-quiz scores, feeling-of-knowing ratings, and hindsight bias all increased with ranked expertise. The fact that there was a hindsight bias at all (hindsight bias for the lowest ranked domain was significantly greater than zero, $t(76) = 12.99, p < .01$) indicates that participants reported that they would have known more than they actually would have. But, why should this overestimation of what one would have known increase with subjective relative

expertise? Werth and Strack (2003) suggested that the knew-it-all-along effect is related to individuals' experienced feelings of confidence when they encounter a question and its answer, because confidence provides individuals with a sense that they already knew the answer to the question. Furthermore, Bradley (1981) showed that confidence increases with ranked expertise. Together these two findings provide a straightforward link between ranked expertise and the knew-it-all-along effect. But, maybe the relationship between these variables is slightly more complex. While Werth and Strack (2003) argued that confidence gives individuals a feeling that they already knew an answer, the authors did not explain what leads to the increased confidence in the first place. The current experiment suggests that participants' familiarity with a domain gives them a feeling that they know the answer to the question (as evidenced by the relationship between rank and FOK ratings). However, an increased *feeling* that one knows the answer to a question does not mean that an individual *actually* knows the answer to the question; participants did not answer many questions correctly, even in their highest ranked domain. The disconnect between what individuals feel like they would have known and what they actually knew is what defines hindsight bias. The link between feeling of knowing and hindsight bias demonstrated in the current experiment has not been explored in previous literature, but it provides evidence for the mechanism underlying hindsight bias, namely a feeling of knowing.

CHAPTER 6

General Discussion

Research on hindsight bias has demonstrated that it is a robust effect with consequences for novices and experts alike (e.g., Arkes et al., 1981; Berlin, 2000; Dawson et al., 1988; Leary, 1982). In the current series of experiments, I explored the relationship between expertise and hindsight bias and found that it is rather complex. Common sense might suggest that experts should be “better” than non-experts, regardless of the task, and this sense of expert superiority is often valid. Everyone would choose to have their appendix removed by a knowledgeable physician than by their hair dresser and their car fixed by a professional mechanic than by a federal judge. However, there are tasks for which expertise may afford no benefits and may even prove to be a detriment (see Shanteau, 1992, for a description of tasks for which expertise is typically beneficial or detrimental). While many investors pay stock brokers large commissions for their expert advice, those who are more expert in finance have been shown to be less accurate in their stock forecasts (Yates et al., 1991). Physicians who have had more experience performing heart catheterizations are more overconfident in their predictions of estimates of values relevant to this procedure (Dawson et al., 1993). The three experiments reported in this paper shed light on the complicated relationship between expertise and

hindsight bias. Those possessing extremely low or extremely high levels of expertise appear to be protected from falling prey to the hindsight bias, while those with intermediate levels of expertise display the bias most readily.

There are many difficulties involved in studying the effects of expertise on hindsight bias. First, expertise is broadly defined, if at all (Shanteau, 1992). Expertise can be a relative or absolute concept. For example, some researchers define expertise based on the number of hours individuals have spent practicing a certain task (Ericsson et al., 1993), while others categorize expertise based on the number classes participants have completed (Glenberg & Epstein, 1987) or the number of years they have spent in a particular profession (Dawson et al., 1993). Expertise can also mean different things in different contexts and amongst different groups of people. For example, while physicians might all be considered experts in comparison to laypeople, physicians can vary in expertise amongst themselves (Dawson et al., 1988). All of these issues regarding the definition of expertise complicate the comparison of experimental results between studies found in the literature. Furthermore, these differences in the categorization of experts are not inconsequential and may partially be responsible for the lack of consensus regarding the effects of expertise on certain biases.

Shanteau (1992) attempted to identify more specifically different levels of expertise, noting that there may actually be three levels of expertise: naïve, novice, and expert. Naïve individuals have little to no knowledge of or experience with a particular domain, novices have intermediate levels of knowledge and experience (and may even perform at “sub-expert” levels), and experts are recognized as performing at the highest levels by members of their professions. While Shanteau’s (1992) labels may be just as

arbitrary as any, they at least highlight the fact there is more differentiation among individuals than just expert and non-expert. When I began the current research on the effects of expertise on the hindsight bias, my intention was to explain the mechanism underlying hindsight bias, namely feeling of knowing. I posited that hindsight bias would increase with expertise because feeling of knowing should increase with (subjective) expertise; the more individuals *feel like* they know about a particular domain, the more they should feel like they knew all along the answers to questions within this domain. I did not, however, make fine-grained distinctions between variations amongst experts and non-experts. But, after analyzing the data from the Ohio State University baseball team (Experiment 2), it became clear that where individuals fall on the expertise continuum actually makes a difference in their performance in hindsight tasks. Therefore, this series of experiments makes two major contributions to the existing hindsight bias literature. First, the results suggest that feeling of knowing may be responsible for hindsight bias. Second, they support an inverted U-shaped relationship between expertise and hindsight bias.

It follows easily from a feeling-of-knowing explanation of hindsight bias why naïve participants would not fall prey to the bias: they are unfamiliar with everything having to do with the domain, so a feeling of *not* knowing (Glucksberg & McCloskey, 1981; Klin, Guzman, & Levine, 1997) arises and hindsight bias is avoided. However, if feelings of knowing are responsible for hindsight bias, why did the varsity baseball players in Experiment 2 fail to demonstrate the bias? Surely these most expert participants were familiar with the people, teams, and events referred to in the baseball questions, so why did familiarity not engender feelings of knowing and consequently

hindsight bias in these participants? One explanation, which has received mixed support at best (e.g., Lichtenstein & Fischhoff, 1977), is Aristotle's explanation: "The more you know, the more you know you don't know." However, there is also a feeling-of-knowing explanation for the lack of hindsight bias in experts. Hoch and Lowenstein (1989) argued that outcome information can produce an "I never would have known it" response (similar to a feeling of *not* knowing) under two conditions: 1) if individuals are completely unfamiliar with the outcome and 2) if individuals are confident about an outcome that turns out to be incorrect. As I already noted, the first case can be easily explained with a feeling-of-knowing explanation of hindsight bias. Similarly, a feeling-of-knowing account of hindsight bias can explain why true experts would not display hindsight bias, though the explanation is more complicated. When the baseball players read each question, they likely generated possible answers. Additionally, since the baseball players were highly familiar with baseball information, they should have been very confident in these initial answers (Bradley, 1981; Dawson et al., 1993). According to Hoch and Lowenstein (1989), then, when the baseball players saw a correct answer that was different from their initial response, the discrepancy should have produced an "I never would have known that" response. As Hoch and Lowenstein (1989) suggested, the surprise that accompanies outcome feedback "acts as an antidote to normal overconfidence" (p. 607).

While the procedure employed in Experiment 2 does not allow for definitive evidence that the baseball players thought of answers to each question before seeing the correct answer, it seems reasonable to assume that they did so spontaneously. Future studies could include a task in which participants answer each question and are then

provided with feedback about the correct answer. The effect of surprise would likely be greater in such an experiment (i.e., surprise would lead to a greater reduction in hindsight bias), but hindsight bias could not be reduced much more than it was in Experiment 2 (average hindsight bias for the baseball players was -2.5, which is extremely close to 0). Future studies with extreme experts could also include a rapid feeling-of-knowing rating portion, which was not included in the experiment using baseball players here. If baseball players gave high FOK ratings for all of the questions (since they are familiar with the information in the questions) but also gave accurate hindsight estimates, this would be strong evidence for the notion that true experts experience an “I never would have known that” response after receiving outcome information that disputes their prior beliefs.

It should be noted here that the effects of surprise on the hindsight bias have been debated in the literature. Ofir and Mazursky (1997) and Mazursky and Ofir (1990) have argued that surprising outcomes can eliminate or reverse the hindsight bias, while Pezzo (2003) and Schkade and Kilbourne (1991) have contended that surprising outcomes can actually “reinforce” or exacerbate the bias. Accounts supporting the attenuating role of surprise on the hindsight bias suggest that a highly surprising outcome triggers an “I could not have known that” feeling and a subsequent effortful search for reasons that could explain the outcome. The great effort involved in uncovering such reasons is thought to confirm the notion that the decision maker could not have predicted the surprising outcome. On the other hand, the same sense-making mechanism thought to reduce the hindsight bias (e.g. Ofir & Mazursky, 1997) has also been used to explain how surprise can lead to greater hindsight bias (e.g., Pezzo, 2003). Pezzo (2003) argued that

when a surprising outcome activates sense-making, a successful resolution of the initially surprising outcome leads to hindsight bias. That is, if the initially surprising outcome can be reconciled easily during the sense-making process, individuals will feel that the outcome really was foreseeable. While experimental evidence exists for both sides of the argument, neither line of research can be clearly applied to the current studies. The sense-making process described by the abovementioned researchers is presumed to occur after a surprising event, where there are multiple antecedents that can be made sense of or not. For example, Pezzo (2003) studied the effects of surprise on hindsight bias after participants were or were not surprised at the outcome of a basketball game, and Ofir and Mazursky's (1997) participants read a scenario about a man who underwent heart surgery for a heart condition and either did or did not die from the surgery. Unlike the types of materials used in the aforementioned experiments, the materials employed in the current series of experiments were less conducive to sense-making, since they were trivia-type questions. Therefore, rather than relying on the sense-making explanation, I suggest that Hoch and Lowenstein's (1989) studies are more relevant to the current line of research, since these authors used trivia-type questions in which sense-making is less likely to play a role. Hoch and Lowenstein (1989) suggested that individuals use "internal surprise reactions" (p. 618) to gauge how likely it is that they would have known the answer to a particular question; this describes the mechanism that I have argued eliminated the hindsight bias for the baseball players in Experiment 2.

Reasons for a lack of hindsight bias in individuals falling at the two extremes of the theoretical expertise continuum have been presented. In terms of the inverted U-shaped relationship that I have suggested exists between expertise and hindsight bias,

those who lie somewhere between these two endpoints are expected to show the greatest amount of hindsight bias. Such individuals know “just enough to be dangerous.” That is, they are familiar enough with the terminology in the questions to produce feelings of knowing, but they are not so expert that answers different from the ones they spontaneously considered would induce surprise and consequently an “I never would have known that” response. Participants in the current Experiment 3 likely knew “just enough to be dangerous.” In fact, the questions presented to participants in Experiment 3 were developed specifically to elicit feelings of knowing without concomitant accuracy. The questions in Experiment 3 (see Appendix D) were developed with the intention that participants would be familiar with the terminology in the questions (i.e., they contained familiar memory “pointers”), but that they would not necessarily be able to produce the answers when prompted to do so. Note, however, that these were not “trick” questions like ones often used in hindsight bias studies (Christensen-Szalanski & Willham, 1991), nor were they “impossible” (e.g., Bradley, 1981). Instead, they were questions dealing with information that typical students would have encountered at some point during their schooling, and they were not particularly difficult or obscure. The finding that participants displayed a significant hindsight bias even in the domain they ranked lowest in expertise suggests that college undergraduates knew enough about these questions to encounter feelings of knowing with these questions in general.

While participants in Experiment 3 should have been familiar with the information in the questions, they nevertheless displayed poor performance when answering the questions. This poor performance suggests that these participants fell somewhere left of center (i.e., they possessed less than intermediate knowledge about the

questions in any domain) in the inverted U-shaped distribution relating expertise to hindsight bias (see Figure 7). “Left of center” is, of course, an arbitrary region on the hypothesized distribution, since expertise is not clearly or objectively defined in this case. However, if we accept the notion that participants in Experiment 3 fell somewhere between extreme naiveté and the maximum point on the distribution, then these participants’ hindsight bias values follow remarkably well from what would be expected from the shape of the distribution. The results of Experiment 3 demonstrated that hindsight bias increased monotonically with expertise, which is precisely what would be expected for individuals with little to just-less-than-moderate knowledge.

The results of the three experiments presented here provide strong evidence for a feeling-of-knowing account of hindsight bias; feelings of *not* knowing can explain why experts and naïve individuals do not display hindsight bias, while individuals with intermediate levels of knowledge show greater hindsight bias as expertise increases because feeling of knowing increases. However, there are some limitations to the current series of experiments that should be addressed. First, the proposed inverted U-shaped relationship between expertise and hindsight bias involves the integration of results across all three experiments. Participants’ levels of expertise in Experiment 1 ranged from naiveté to novice. In this experiment, hindsight bias increased as a function of expertise, which is consistent with the inverted U-shaped relationship between expertise and hindsight bias. In Experiment 2, levels of expertise also ranged from naiveté to novice for those considered non-experts, but true experts were also included. In the second experiment, the non-experts as a group displayed hindsight bias, while the experts showed virtually no hindsight bias at all. Again, this is consistent with an inverted U-

shaped expertise-hindsight bias relationship. Finally, participants in Experiment 3 were also assumed to have novice levels of expertise, and in this sample hindsight bias again increased with expertise, as predicted by the inverted U-shaped relationship. While the results of each experiment are independently consistent with the proposed relationship between expertise and hindsight bias, the argument for this relationship would be stronger if the relationship were observed within the same experiment. For example, if middle school students, undergraduate biology majors, first-year medical students, residents, and senior physicians were all asked the same questions about medical facts, the inverted U-shaped relationship between expertise and hindsight bias could be tested within the same study. Middle school students and senior physicians would be expected to display no hindsight bias, undergraduate biology majors and first-year medical students would likely demonstrate the greatest levels of hindsight bias, and residents would likely display lower levels of hindsight bias than the biology majors and first-year medical students.

While the abovementioned predictions follow from the proposed expertise-hindsight bias relationship, they also highlight a second limitation of the current studies. None of the experiments presented here demonstrates the relationship between expertise and hindsight bias to the “right of center” on the U-shaped curve. This limitation is due to the unanticipated results of Experiment 2. As mentioned in Chapter 4, the result showing no hindsight bias in the baseball player sample was unexpected, as I had predicted a monotonically increasing linear relationship between expertise and hindsight bias. Therefore, no experiment was included in the current series that specifically addresses what happens to hindsight bias between intermediate levels of expertise and

“true” expertise. The entire right-hand side of the distribution is not explored here. It is entirely possible that hindsight bias is absent for naïve individuals, increases monotonically with expertise, and disappears only for true experts. Such a relationship would also be supported by the results in the current three experiments. The study proposed above including participants spanning the entire range of medical expertise would help to answer more definitively what the true shape of the distribution is likely to be.

A final limitation of the current series of experiments is that feeling of knowing was only measured directly in Experiment 3. While the proposed relationship between familiarity, feeling of knowing, and hindsight bias can account for all of the results in each experiment, FOK ratings were not obtained in Experiments 1 and 2. Collecting FOK judgments from the baseball players would have been particularly informative, since the nature of the predicted relationship between feeling of knowing and hindsight bias is expected to be different for true experts and less-knowledgeable individuals. While previous research (Hoch & Lowenstein, 1989) supports the notion that the baseball players experienced an “I never would have known that” effect when appropriate, thereby eliminating hindsight bias, this proposition is speculative at this point.

Let us return to the examples of Mary and Professor M presented earlier in this paper and include as well Mary’s daughter Ally, who is in 5th grade. Imagine they were playing a trivia game similar to Jeopardy!, and the categories “Hannah Montana,” “Principles of Factor Analysis,” and “Around the Gourmet Market” were available. Note that the Jeopardy! paradigm is similar to the feeling-of-knowing procedure employed in Experiment 3, since Jeopardy! contestants are given only a few seconds to respond once

Alex Trebek asks the question. In light of the findings from the three experiments presented above, when would we expect each contestant to buzz in with an answer, and how would we expect them to feel once the answer is revealed? Since neither Mary nor Ally has ever heard of factor analysis, we should not expect them to even attempt to answer any of the questions in “Principles of Factor Analysis.” These questions do not produce feelings of knowing for Mary or Ally, since they are completely unfamiliar with the terminology in the questions. Furthermore, when the answers to these questions are revealed, Mary and Ally should experience an “I never would have known that” response. Ally, who has seen every Hannah Montana episode and reads everything she can about the stars of the show, would likely buzz in often for questions from the “Hannah Montana” category, since the people and events mentioned in the question are very familiar to her. However, if Ally gives an answer that is incorrect, she will likely be surprised at the correct answer, because she was confident in her initial response. Buzzing in is, in some senses, a proxy for confidence, since contestants are penalized for incorrect answers. When Ally hears the answer, which is different from her initial response, a feeling of “I never would have known that” is likely to arise. A similar chain of events occurs for Professor M when questions from the “Principles of Factor Analysis” category are asked. Mary should have a hard time with two of the categories: “Hannah Montana” and “Around the Gourmet Market.” Mary has some knowledge about both of these categories (she has been exposed to Hannah Montana through her daughter and she is an avid cook), though she knows more about cooking than she does Hannah Montana. When Mary hears questions from either of these categories, she is likely to buzz in and attempt to answer them, since she is familiar with the terminology in the questions,

producing feelings of knowing. If Mary is wrong, she will experience a “Darn, I knew that!” reaction, not an “I never would have known that” reaction, because she is not so expert in either category to be extremely confident. Though these reactions cannot be quantified in this example, Mary’s “Darn, I knew that!” reaction should be greater in the “Around the Gourmet Market” category than in the “Hannah Montana” category (but, she is also less likely to attempt to answer the Hannah Montana questions than the food-related questions if buzzing in is a proxy for confidence).

The three experiments presented herein contribute significantly to the existing literatures in both expertise and hindsight bias. While the question of how to define expertise remains largely unanswered, the current series of experiments demonstrates that differences in levels of expertise are meaningful and can result in important variations in individuals’ responses to certain stimuli. Variations in expertise were shown in three experiments to significantly affect the degree to which individuals displayed hindsight bias. Underlying these differences in levels of hindsight bias is a difference in the way individuals with varying levels of expertise process familiar and unfamiliar information. In the current series of experiments, feelings of knowing or *not* knowing, resulting from participants’ familiarity with a particular domain, were shown to be responsible for the way individuals deal with outcome information in light of their prior knowledge. While completely naïve individuals and “super experts” appear to be protected from hindsight bias, those with levels of expertise falling somewhere between these two extremes are susceptible to the bias, though to varying degrees. This inverted U-shaped relationship between expertise and hindsight bias poses a dilemma for most communications taking place between more-informed and less-informed individuals. While physicians,

professors, and other professionals regularly are sought out for their expertise in their respective fields, most professionals are likely to be sub-expert (i.e., not every physician, professor, mechanic, or pilot can be considered to be at the very top of his field). Hindsight bias, therefore, is likely to run rampant in most populations, making communication with less-informed others unsatisfactory and difficult.

LIST OF REFERENCES

- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68, 804-825.
- Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, 73, 305-307.
- Arkes, H. R., & Freedman, M. R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition*, 12, 84-89.
- Arkes, H. R., & Harkness, A. R. (1980). The effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 568-575.
- Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, 66, 252-254.
- Benner, P. (1982). From novice to expert. *The American Journal of Nursing*, 82, 402-407.
- Berlin, L. (2000). Hindsight bias. *American Journal of Roentgenology*, 175, 597-601.
- Bernstein, D. M., & Harley, E. M. (2007). Fluency misattribution and visual hindsight bias. *Memory*, 15, 548-560.
- Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, 11, 491-504.
- Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society*, 17(2), 82-84.

- Bruner, J. S. (1957). Going beyond the information given. In H. Gruber, K. Hammond, & R. Jessor (Eds.), *Contemporary approaches to education*. Cambridge, Mass: Harvard University Press.
- Camerer, C. F., & Johnson, E. J. (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In W. M. Goldstein, & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections, and controversies* (pp. 342-364). New York, NY: Cambridge University Press.
- Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, 51, 605-620.
- Caplan, R. A., Posner, K. L., & Cheney, F. W. (1991). Effect of outcome on physician judgments of appropriateness of care. *Journal of the American Medical Association*, 265, 1957-1960.
- Casper, J. D., Benedict, K., & Kelly, J. R. (1988). Cognitions, attitudes, and decision-making in search and seizure cases. *Journal of Applied Social Psychology*, 18, 93-113.
- Casscells, W., Schoenberger, A., & Graboys, T. (1978) Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999-1001.
- Chapman, G. B., & Elstein, A. S. (2000). Cognitive processes and biases in medical decision making. In G. B. Chapman & F. A. Sonnenberg (Eds.), *Decision Making in Health Care: Theory, Psychology, and Applications* (pp. 183-210). New York, NY: Cambridge University Press
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, 13, 257-273.
- Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48, 175-168.
- Costermans, J. Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 142-150.
- Dawson, N. V., Arkes, H. R., Siciliano, C., Blinkhorn, R., Lakshmanan, M., & Petrelli, M. (1988). Hindsight bias: An impediment to accurate probability estimation in clinicopathologic conferences. *Medical Decision Making*, 8, 259-264.

- Dawson, N. V., Connors, A. F., Speroff, T., Kemka, A., Shaw, P., & Arkes, H. (1993). Hemodynamic assessment in managing the critically ill: Is physician confidence warranted? *Medical Decision Making*, 13, 258-266.
- de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton. (Original work published in Dutch, 1946)
- Dreyfus, S. E. (2004). The five stage model of adult skill acquisition. *Bulletin of Science, Technology, & Society*, 24, 177-181.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49, 725-747.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363-406.
- Fischhoff, B. (1975). Hindsight $\neq$ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422-444). New York: Cambridge University Press.
- Fischhoff, B. & Beyth, R. (1975). "I knew it would happen"—Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, 13, 1-16.
- Gladwell, M. (2008). *Outliers: The Story of Success*. New York: Little Brown & Company.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15, 84-93.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119-136.
- Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 311-325.
- Gray, R., Beilock, S. L., & Carr, T. H. (2007). "As soon as the bat met the ball, I knew it was gone": Outcome prediction, hindsight bias, and the representation and control

- of action in expert and novice baseball players. *Psychonomic Bulletin & Review*, 14, 669-675.
- Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, 26, 103-117.
- Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review*, 86, 777-830.
- Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review*, 93, 1-44.
- Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, 25, 48-63.
- Harley, E. M., Carlsen, K. A., & Loftus, G. R. (2004). The "saw-it-all-along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 960-968.
- Hastie, R., Schkade, D. A., & Payne, J. W. (1999). Juror judgments in civil cases: Hindsight effects on judgments of liability for punitive damages. *Law and Human Behavior*, 23, 597-614.
- Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied*, 5, 205-221.
- Hoch, S. J., & Lowenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 605-619.
- Kern, L., & Doherty, M. E. (1982). "Pseudodiagnosticity" in an idealized medical problem-solving environment. *Journal of Medical Education*, 57, 100-104.
- Klin, C. M., Guzman, A. E., & Levine, W. H. (1997). Knowing that you don't know: Metamemory and discourse processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1378-1393..
- Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 686-715). Cambridge: Cambridge University Press.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609-639.

- Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124, 311-333.
- Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 34-53.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica*, 41, 151-164.
- LaBine, S. J., & LaBine, G. (1996). Determinations of negligence and the hindsight bias. *Law and Human Behavior*, 20, 501-516.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavioral and Human Performance*, 20, 159-183.
- Leary, M. R. (1982). Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin*, 8, 257-263.
- Marks, M. A. Z., & Arkes, H. R. (2009, in press). The effects of mental contamination on the hindsight bias: Source confusion determines success in disregarding knowledge. *Journal of Behavioral Decision Making*.
- Mazursky, D., & Ofir, C. (1990). I could never have expected it to happen: The reversal of the hindsight bias. *Organizational Behavior and Human Decision Processes*, 46, 20-33.
- McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, 107, 179-191.
- McNeil, B. J., Pauker, S., Sox, H., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306, 1259-1262.
- Metcalf, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 288-294.

- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 851-864.
- Musch, J., & Wagner, T. (2007). Did everybody know it all along? A review of individual differences in hindsight bias. *Social Cognition*, 25, 64-82.
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19, 338-368.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 26, pp. 125-173). San Diego, CA: Academic Press.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737-759.
- Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologica*, 64, 245-259.
- Ofir, C., & Mazursky, D. (1997). Does a surprising outcome reinforce or reverse the hindsight bias? *Organizational Behavior and Human Decision Processes*, 69, 51-57.
- Pezzo, M. V. (2003). Surprise, defense, or making sense: What removes hindsight bias? *Memory*, 11, 421-441.
- Pohl, R. F. (1992). Der Rückschau-Fehler: systematische Verfälschung der Erinnerung bei Experten und Novizen [Hindsight bias: Systematic distortions of the memory of experts and laymen]. *Kognitionswissenschaft*, 3, 38-44.
- Pohl, R. F., Schwarz, S., Sczesny, S., & Stahlberg, D. (2003). Hindsight bias in gustatory judgments. *Experimental Psychology*, 50, 107-115.
- Poses, R. M., & Anthony, M. (1991). Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia. *Medical Decision Making*, 11, 159-168.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90-138.
- Reder, L. M. (1988). Strategic control of retrieval strategies. *The Psychology of Learning and Motivation*, 22, 227-259.

- Redelmeier, D. A., & Shafir, E. (1995). Medical decision making in situations that offer multiple alternatives, *Journal of the American Medical Association*, 273, 302-305.
- Schkade, D. A., & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes*, 49, 105-123.
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, 252-266.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61, 394-403.
- Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism. *Psychological Bulletin*, 109, 472-489.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-290.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, S251-S278.
- Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory*, 11, 411-419.
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117-142.
- Winman, A., Juslin, P., & Bjorkman, M. (1998). The confidence-hindsight mirror effect in judgment: An accuracy-assessment model for the knew-it-all-along phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 415-431.
- Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, 49, 60-79.

APPENDIX A
Baseball and Dogs Essays and Baseball Quiz

Baseball Essay

The 2006 Most Valuable Player awards in the American and National Leagues were among the most unusual in baseball history. In the American League, Justin Morneau won the award. Morneau was benched in June with a terrible .236 batting average. He was not selected to the July All-Star game due to his mediocre play, only the tenth player in major league history to win the MVP Award in a year despite not even appearing in the All-Star game that year. However after being benched and then returning to the line-up, Morneau batted an incredible .362 with 92 RBIs during the last three and a half months of the season. Another unusual feature of Morneau's award is that Morneau is Canadian, only the second Canadian player in baseball history ever to receive the MVP Award, the other being Larry Walker of the Colorado Rockies in 1997. Yet another bizarre feature of Morneau's award is that he was one of the lowest paid players in the American League in 2006. And finally, Morneau did not come even close to being the best hitter on his team, the Minnesota Twins. Catcher Joe Mauer of the Twins hit over 25 points higher than Morneau but finished sixth in the MVP voting. Morneau edged out Derek Jeter of the Yankees, who had another fine year. Jeter batted

.344. Boston's David Ortiz finished only third in the MVP balloting, even though he led the American League with 137 RBIs and 54 home runs. The last Minnesota Twin to win the MVP award was the great Rod Carew, who won in 1977. Carew finished his career with over 3,000 hits. He was the only player other than Ty Cobb to win the batting crown for three straight years. Like Morneau, Carew was foreign-born, having come to the United States from Panama. Before Carew the previous Twin to win the award was Harmon Killebrew in 1969. At the time of his retirement, Killebrew had hit more home runs than any other right-handed batter in American League history.

In the National League the MVP winner in 2006 was the Phillies' Ryan Howard. He was the first Phillie to win the award since Ohio-born Mike Schmidt won it twenty years earlier. Howard was one of the lowest-paid players in the National League. Howard didn't even get a chance to play in the major leagues until the middle of the prior year (2005) when Phillies first-baseman Jim Thome went on the disabled list. Howard won the Rookie of the Year award that year, and by winning the MVP award in 2006, Howard became only the second player in baseball history to win the Rookie of the Year award and MVP award in consecutive years, the other being Cal Ripken of the Orioles. In 2006 Howard had 58 homers, the most in the major leagues in five years—Barry Bonds having hit 73 in 2001. The person who finished second in the NL voting was Albert Pujols. It is ironic that during the winter Pujols and Howard often worked out together at the same facility near St. Louis, and Pujols gave Howard important tips on how to improve his batting. By giving Howard this advice, Pujols may have cost himself the MVP award. By finishing second in the MVP voting, Pujols joined Mickey Mantle as the only player to finish second three times. During a game against the Yankees in June,

Howard was the first player to hit a home run into the third deck of Citizens Bank Park in Philadelphia. Finishing third in the NL MVP voting was Lance Berkman of the Astros, who started the season on the disabled list after hurting himself during the winter.

Dogs Essay

Owning a dog can be the beginning of years of happiness as the special bond between humans and canines exceeds even the greatest of expectations. However, to ensure the best relationship with your dog, you must be prepared for some important responsibilities, including whether or not you have found the right breed of dog to fit into your lifestyle and home, whether or not you will have enough time to spend training, grooming and exercising your dog, keeping in mind that letting a dog out just to go to the bathroom does not constitute adequate exercise, and whether or not you are willing to spend the resources to ensure the best future for your dog. Whether there is a breed of dog that you've had your eye on or whether you're confused about how to select a dog, you should do some homework to make sure that you select the right dog for you and your family.

The benefit of selecting a purebred dog is their predictability in size, coat, care requirements and temperament, and knowing what your cute puppy will look like and the kind of care he will need as an adult is a key in selecting the breed for you. Too frequently, common sense goes out the window when it comes to buying a puppy, especially when the purchase is by a family with children. Buying a dog is like buying anything else; the more you know before you buy, the better off you will be. This advice applies to all aspects of buying your dog, from selecting the breed to deciding where to obtain the puppy.

While investigating, always be honest with yourself, for the Bearded Collie you fell in love with because of his lush coat is indeed beautiful, but are you going to be able to brush this coat every day as it requires? Think about the size of your house or your apartment and whether or not that Golden Retriever will be happy in your studio apartment. The Golden Retriever is a larger sporting dog who requires a lot of exercise, so having a fenced yard so he can go out safely is important to think about, keeping in mind that if your yard is not currently fenced, doing so may be an expensive endeavor. These are crucial questions regarding the safety of your dog and being a responsible neighbor. Always remember, it is okay to change your mind about which breed you want or if you want the responsibility of owning a dog at all.

Buying your puppy from a responsible and well-respected breeder is a recommendation that cannot be stressed enough because responsible breeders are concerned with the betterment of the breed. For example, they work on breeding healthier dogs with the appropriate temperament for their breed. Once you select a breeder, screen the breeder and ask to see at least one of the parents (the dam or the sire) of your puppy. See how the dogs in your breeder's home interact with your breeder; are they friendly and outgoing or do they shy away? The responsible breeder will be screening you, too, looking for the best home for each puppy. This is not the time to hunt for a bargain. Your new puppy will be a member of your family for his lifetime, so you'll want to make a wise investment, keeping in mind that the purchase price of your puppy is not the only cost you have to consider. Be aware that the puppy you bring home will need proper care: food, health care, (a dog needs annual shots). Evaluate your budget; ask yourself if you really can afford a dog.

20-Question True/False Quiz

1. Harmon Killebrew was the Twin who won the MVP award just prior to Morneau. (F)
2. Howard received batting tips from Jim Thome (F)
3. Howard hit a monster home run in Yankee Stadium (F)
4. Albert Pujols finished second in the National league MVP balloting in 2006 (T)
5. Ryan Howard didn't play in the major leagues during the first half of 2005. (T)
6. David Ortiz finished second in the American League MVP voting in 2006. (F)
7. Larry Walker won the MVP award while playing for the Minnesota Twins. (F)
8. In 2001 Barry Bonds hit 73 home runs. (T)
9. Morneau was injured during the winter and was unable to play until June of 2006. (F)
10. Joe Mauer is a catcher for the Twins. (T)
11. Mike Schmidt played for the Phillies. (T)
12. Rod Carew won the MVP award three times. (F)
13. Harmon Killebrew was a right-handed home run hitter. (T)
14. Mickey Mantle led the league in hitting for three consecutive years. (F)
15. Cal Ripken won both the MVP and Rookie of the Year awards. (T)
16. In 2006 Howard hit more home runs than anyone else in the National League had hit in any of the prior five years. (T)
17. Morneau won the Rookie of the Year award in 2005. (F)
18. Howard was born in Panama. (F)
19. Larry Walker was born in Canada. (T)
20. Mike Schmidt was born in Ohio. (T)

APPENDIX B
Baseball Expertise Quiz

1. If the visiting team is ahead, how many innings must a game go to be official?
2. What is the official score of a forfeited game?
3. If a batter makes an out in order to advance a runner one base, the play is termed a(n):
4. What does RBI stand for?
5. During a game a pitcher warms up before going into the game. The place in which he warms up is called the:
6. When a runner on base starts to run with the pitch and the batter knows he is going to run and tries definitely to hit the ball, this is termed:
7. Trying to get a runner from third base to home plate by bunting the ball is called a(n):
8. The distance from the pitcher's mound to home plate is:
9. The distance between any two bases is:
10. A pitcher who does not start a game but comes into a game at some point after it has started is called a(n) _____ pitcher.
11. If during a game a manager objects to the interpretation of a rule given by an official of the game, the manager may play the game _____.
(what is the expression that comes next?)

12. How many innings must a starting pitcher pitch in order to receive credit for a win?
13. When a pitcher makes an illegal motion with a runner on base, the runner gets to advance a base. The advancing of a base due to the pitcher's illegal action is called a(n):
14. On each side of home plate there are rectangular markings where a batter stands when he is hitting. The marked-off area where the batter stands when he is hitting is the:
15. When a team is at bat, one person of that team stands in a rectangular area by first base and another person of that team stands in a rectangular area near third base. These two rectangular areas are called:
16. With a man on first base the pitcher throws the ball and the catcher lets the ball get by him, allowing the runner to go from first to second base. The official charges the _____ with a(n) _____.
17. This time the man on first base gets to second base because the pitcher threw the ball behind the batter and the catcher was unable to stop it. In this case the official scorer charges the _____ with a(n) _____.
18. An infielder goes "deep into the hole" and gets the batter out at first base. Which infielder did this?
19. If a ground ball hit by the batter rolls foul and then rolls fair before the ball reaches third base, is the ball fair or foul?
20. When the pitcher suddenly turns and throws to second base to try to make a runner out who is on second base, this action is called:
21. If a batter is hit by a pitched ball, what do the rules say he is entitled to do?
22. What does ERA stand for?

23. A batter who is scheduled to bat immediately after the batter currently hitting is referred to as being:
24. In the American League, what does dh stand for?
25. A batter who hits either right- or left-handed is called a(n):
26. A batter hits a ground ball to the third baseman, who, in throwing the ball to first base, throws the ball over the head of the first baseman and into the stands. Is the batter safe in first, safe on second, safe on third, or does he get to score?
27. When the championship team of the American League plays the championship team of the National League, it is called the World Series. What is the number of winning games for one team it takes to win the World Series?
28. A player who has not been in a game is put up to bat to hit for someone who has been in the game. The player put into the game is called a(n):
29. In baseball games, one person determines whether a pitch is a ball or a strike. This person is a(n):
30. When is a batter allowed to run to first base, even though he struck out?
31. With runners on second and third base, a batter may be walked intentionally. Assume the batter is not an outstanding hitter. What would be the major purpose of such a move?
32. What type of pitcher generally has the best “move” to first base when trying to pick off runners?
33. Pitchers tend to throw “off-speed” or “change of pace” pitches when facing a _____ hitter.

APPENDIX C
Baseball Current Events Quiz

1. Who was named the 2005 American League Most Valuable Player (MVP)? (Alex Rodriguez)
2. What team won the 2004 American League Championship Series (ALCS)? (Boston Red Sox)
3. Former New York Yankees general manager, Joe Torre, became the manager of what team for the 2008 season? (Los Angeles Dodgers)
4. Who won the 2006 National League Cy Young award? (Brandon Webb)
5. What were the Nationals called before relocating to Washington D.C. in 2005? (Montreal Expos)
6. What famous country singer went to spring training with the San Diego Padres in 1998? (Garth Brooks)
7. What is the name of the player who, in 1994, almost became the first player in 50 seasons to finish the season with a .400 batting average? (Tony Gwynn)
8. Of the Detroit Tigers, Houston Astros, Milwaukee Brewers, and New York Yankees, which team never changed its name during the entire 20th century? (Detroit Tigers)

9. Who was the first player in history to hit 400 home runs and steal 400 bases? (Barry Bonds)
10. During the 20th century, who was the only player to steal 1,000 career bases? (Rickey Henderson)
11. St. Louis Cardinal Jim Edmonds began his major league career with what team? (California Angels)
12. In 2007, what major league baseball team became the first professional sports team to lose 10,000 games? (Philadelphia Phillies)
13. On July 16, 2007, which slugger moved into the number six position for career home runs? (Ken Griffey Jr.)
14. In 2007, what career Houston Astro became the 27th member of the 3000-hit club in Major League baseball? (Craig Biggio)
15. Which Texas Ranger hit his 600th career home run in 2007? (Sammy Sosa)
16. After 16 seasons with the Braves, it took another five with the Mets for which pitcher to become the 25th pitcher in baseball history to win 300 career games? (Tom Glavine)
17. By giving up 22 home runs in 2007, which pitcher moved into the top 10 for the most home runs given up in a career, passing the 400-mark? (David Wells)
18. In just 17 seasons of Major League baseball, which 2007 White Sox player moved into the number three position for most career strikeouts, behind Sammy Sosa and Reggie Jackson? (Jim Thome)
19. Which 41-year old Boston Red Sox pitcher appeared in his 1,000th game in 2007? (Mike Timlin)

20. Which player won the 2007 home run derby with a total of 17 home runs? (Vladimir Guerrero)

APPENDIX D
Five-Topic General Knowledge Quiz

History

1. What is the last name of the man who founded Standard Oil in 1870? (Rockefeller)
2. What Portuguese explorer was the first person to lead an expedition across the Pacific Ocean? (Magellan)
3. What Native American woman accompanied Lewis and Clark in their exploration of the Western United States? (Sacagawea)
4. Who was the leader of the Soviet Union during World War II? (Stalin)
5. What prominent figure of the American Revolution is known for saying “Give me liberty or give me death”? (Patrick Henry)
6. What French term is an economic ideology that means minimal government regulation of the economy? (Laissez-faire)
7. What U.S. senator was known for his role in the targeting of suspected communists and communist sympathizers in the United States in the early 1950’s? (McCarthy)
8. Originally issued in the year 1215 and written in Latin, what document is the foundation of the British parliamentary system? (Magna Carta)

9. The Social Security system was introduced as part of what series of economic programs introduced between 1933 and 1936? (New Deal)
10. On what Japanese city did the U.S. drop the first atomic bomb in 1945? (Hiroshima)
11. Born into slavery in the early 1800's, what woman is best known for having "never lost a passenger" in the "underground railroad"? (Harriet Tubman)
12. Who was President of the United States during the Cuban Missile Crisis? (JFK)
13. Launched in October of 1957, what is the name of the first man-made satellite? (Sputnik)
14. What is the name of the pamphlet that Thomas Paine published in 1776 to urge colonists to seek independence from Britain? (Common Sense)
15. What is the name of the settlement that became infamous for being the site where 918 people died from drinking cyanide-laced Cool-Aid? (Jonestown)
16. Who shot Beatle John Lennon four times in the back at the Dakota hotel on December 8, 1980? (Mark David Chapman)
17. Who was charged with the assassination of President John F. Kennedy? (Lee Harvey Oswald)
18. Who shot President Abraham Lincoln on April 14, 1865? (John Wilkes Booth)
19. Which Spanish conquistador is associated with finding the "Fountain of Youth" in Florida? (Ponce de Leon)
20. What American woman is credited with sewing the first American flag, which incorporated stars representing the first thirteen colonies? (Betsy Ross)

Literature

1. What is the name of the Nathaniel Hawthorne character who was forced to wear the “Scarlet Letter”? (Hester Prynne)
2. What is the pen-name of Samuel Clemens? (Mark Twain)
3. Who wrote the books upon which the movies “The Firm” and “The Pelican Brief” were based? (John Grisham)
4. In what famous novel was the term “Big Brother” first introduced? (1984)
5. What is the name of the main character in the J.D. Salinger novel “The Catcher in the Rye”? (Holden Caulfield)
6. What American playwright won a Pulitzer Prize for penning the drama “A Streetcar Named Desire” in 1948? (Tennessee Williams)
7. “Call me Ishmael” is the famous opening line from what 1851 novel? (Moby Dick)
8. What novelist and short-story writer won a Pulitzer Prize in 1953 for penning “The Old Man and the Sea”? (Ernest Hemingway)
9. What is the last name of the author who wrote “The Great Gatsby”? (Fitzgerald)
10. Who wrote “The Count of Monte Cristo” and “The Three Musketeers”? (Alexander Dumas)
11. “The Wife of Bath’s Tale” and “The Knight’s Tale” are both stories contained in “The Canterbury Tales,” written by what 14th century author? (Chaucer)
12. What fictional Spanish knight attacked windmills thinking they were giants? (Don Quixote)
13. What is the name of the monster with whom Beowulf battles in the Old English heroic poem? (Grendel)

14. What epic poem written by John Milton is about the rebellion of Satan and the fall of Adam and Eve? (Paradise Lost)
15. In what country does Shakespeare's "Macbeth" take place? (Scotland)
16. Ebenezer Scrooge is the main character of a famous novel written by what author in 1843? (Charles Dickens)
17. Atticus and Scout Finch are characters in what Pulitzer Prize-winning Harper Lee novel? (To Kill a Mockingbird)
18. What Arthur Miller play is based on the events surrounding the Salem Witch Trials of 1692? (The Crucible)
19. What Nobel Prize-winning American novelist wrote "Absalom, Absalom!" and "The Sound and the Fury"? (Faulkner)
20. What American science fiction writer is best known for writing "Jurassic Park" and its sequel "The Lost World"? (Michael Crichton)

Science

1. What is the name of the organ that produces insulin? (Pancreas)
2. What are the last names of the two men credited for discovering the double-helix structure of DNA? (Watson and Crick)
3. What is the term for the force resisting the relative lateral motion of two solid surfaces in contact? (Friction)
4. The octopus and squid belong to what class of organism meaning "head-footed"? (Cephalopod)
5. What group of chemical elements that includes helium, neon, and xenon are odorless and colorless under standard conditions? (Noble gases)

6. What type of molecules are often referred to as the “building blocks of protein” and play a variety of roles in metabolism? (Amino acids)
7. What is the name of the whip-like structure extending from certain cells that aids in locomotion? (Flagella)
8. What is the name of the male reproductive organ of a flower? (Stamen)
9. What is the name of the thermodynamic (absolute) temperature scale where absolute zero, the theoretical absence of all thermal energy, is zero? (Kelvin)
10. The stapes, incus, and malleus are bones found in what part of the human body? (Ear)
11. What is the name for the protein in red blood cells that carries oxygen? (Hemoglobin)
12. What is the name for the lowest layer of the Earth’s atmosphere, which is the layer in which we live? (Troposphere)
13. What is the last name of the scientist often referred to as the “father of genetics” for his study of the inheritance of traits in pea plants? (Mendel)
14. What is the standard unit of measure for power and is equivalent to one joule per second? (Watt)
15. Which planet has moons named Phobos and Demos? (Mars)
16. Medical terms referring to what organ contain the prefix “renal”? (Kidney)
17. What pigment found in plants is essential for allowing plants to obtain energy from light? (Chlorophyll)
18. What organic compound is the structural component of the primary cell wall of green plants? (Cellulose)

19. From what part of an insect's body do the legs extend? (Thorax)
20. What is the name of the property of living organisms that regulates their internal environment so as to maintain a stable, constant condition? (Homeostasis)

Math

1. What is the term for any natural number that is not a fraction? (Integer)
2. What mathematical concept is used to describe the behavior of a function as its input either "gets close" to some point or becomes arbitrarily large? (Limit)
3. What is the term for the answer to a multiplication problem? (Product)
4. What mathematical term is defined as $2\pi r$? (Circumference)
5. What is the name of triangle that has no equal sides and no equal angles? (Scalene)
6. What mathematical property of relation states that if a relation holds between a and b and between b and c , then it also exists between a and c ? (Transitive Property)
7. A pair of angles that add up to 180 degrees are called what type of angles?
(Supplementary)
8. If $X = 2$, then $\frac{1}{2}$ is said to be the _____ of X . (Inverse)
9. What is the name of a straight line that "just touches" a curve or circle at a certain point? (Tangent)
10. What is the name for the longest side of a right triangle? (Hypotenuse)
11. What is the term for the answer to a division problem? (Quotient)
12. What mathematical term is represented by an exclamation point (!)? (Factorial)
13. Between any two points a line can be drawn. What can be drawn between any three points? (Plane)

14. What is the mathematical term for an integer that evenly divides a number (n) without leaving a remainder? (Divisor)
15. What is the mathematical term for a number whose squared value is a real number not greater than zero? (Imaginary number)
16. In mathematics the symbol $\sqrt{}$, used to represent a square root, is called what? (Radical)
17. What mathematical term refers to a corner of a polygon or a point where lines meet? (Vertex)
18. What mathematical term refers to a line that a graph approaches but never intersects? (Asymptote)
19. What is the term for two angles in a plane which share a common vertex and a common side but do not overlap? (Adjacent)
20. What Greek letter is used in mathematics to describe a “change in” or a “difference of”? (Delta)

Geography

1. What is the name of the European peninsula that contains Spain and Portugal? (Iberian)
2. What city-state is home to the Vatican? (Vatican City)
3. What is the name of the mausoleum located in Agra, India that was built by Emperor Shah Jahan in memory of his favorite wife, Mumtaz? (Taj Mahal)
4. Through what district in southeast London does the Prime Meridian pass? (Greenwich)
5. In what Spanish city does the “running of the bulls” take place each year? (Pamplona)

6. What body of water connects the Atlantic Ocean to the Mediterranean Sea and separates Spain from Morocco? (Strait of Gibraltar)
7. Mount Everest and K2 are part of which Asian mountain range? (Himalayas)
8. What national park established by the U.S. Congress in 1872 is located primarily in Wyoming and extends into Montana and Idaho? (Yellowstone)
9. What U.S. river flows through the Grand Canyon? (Colorado)
10. What is the world's largest island by area that is not a continent? (Greenland)
11. In what U.S. state does the Mississippi river begin? (Minnesota)
12. What is the longest river in the United States? (Missouri)
13. What U.S. state is home to Mount Rushmore? (South Dakota)
14. The majority of the Amazon rainforest is contained in what country? (Brazil)
15. What largest Asian desert covers parts of northern and northwestern China and southern Mongolia? (Gobi)
16. What is the largest country in the world by area? (Russia)
17. In what Canadian province is Toronto located? (Ontario)
18. What is the capital of Australia? (Canberra)
19. What is the capital of the state of Washington? (Olympia)
20. What is the capital of Norway? (Oslo)

APPENDIX E
Tables and Figures

	<i>b</i>	SE
Constant	9.915*	.528
Expertise Quiz Score	.033	.031
Discount	1.225	.672
Do Your Best	.570	.717
Expertise*Discount	.127*	.041
Expertise*Do Your Best	.224*	.043

Table 1: Regression coefficients from Experiment 1

* Values significant at the .01 level

Figure 1: Expertise quiz score results for Experiment 1

Figure 2: Regression Graph for Experiment 1

Figure 3: Expertise quiz score results for Experiment 2

Figure 4: Relationship between expertise and feeling-of-knowing in Experiment 3

Figure 5: Relationship between expertise and questions answered correctly in each domain in Experiment 3

Figure 6: Relationship between expertise and hindsight bias in Experiment 3

Figure 7: Inverted U-shaped relationship between expertise and hindsight bias
*The shaded region is where participants in Experiment 3 are proposed to fall in the distribution.