

Hindsight and Numeric Confidence: Influences on Decision Processes

by

Michael C. Silverstein

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Psychology

Dissertation Committee:

Ellen Peters, PhD, Chair

Paul Slovic, PhD, Core Member

Robert Mauro, PhD, Core Member

John Clithero, PhD, Institutional Representative

University of Oregon

Spring 2025

© 2025 Michael C. Silverstein
This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International license



DISSERTATION ABSTRACT

Michael C. Silverstein

Doctor of Philosophy in Psychology

Title: Hindsight and Numeric Confidence: Influences on Decision Processes

Hindsight bias has long been speculated to increase confidence for subsequent decisions and judgments (e.g., Dawson et al., 1993; Hawkins & Hastie, 1990; Roese & Vohs, 2012); however, to our knowledge, there is no empirical evidence for this consequence of hindsight. The goals of the present dissertation are to: 1) provide evidence of the causal relationship between hindsight bias and increases in both prospective and retrospective confidence, 2) study how the amount of hindsight experience can affect changes in confidence, 3) begin to understand the consequences of elevated confidence due to hindsight bias on decision processes, and 4) investigate possible emotional processes underlying how people learn confidence from experiences. To accomplish these goals, we developed a paradigm that would afford some participants hindsight bias during training for a novel math choice task (Pilot Studies). Then participants completed a set of test trials during which they indicated their confidence in their answers.

Using an online sample of participants (N = 842 MTurkers; Study 1), we found that participants who experienced hindsight trials during training reported greater prospective confidence than participants who completed the training in foresight without any feedback (but not more confidence than those with feedback). Additionally, participants who experienced hindsight trials during training reported greater retrospective confidence than participants who completed the training in foresight without any feedback. However, foresight participants appear to have decreased in prospective confidence, rather than hindsight participants increasing in

prospective confidence, challenging the speculation that hindsight bias increases confidence. Moreover, the amount of training did not appear to play a significant role in participants' confidence judgments. Prospective confidence had a parabolic relation with response times (such that people with moderate levels of prospective confidence answered more slowly), suggesting that prospective confidence modulates decision processes.

Using an in-person sample of undergraduates ($N = 199$; Study 2), we replicated key findings from Study 1 and investigated the role of affective responses to feedback (i.e., participants learning that they were correct/incorrect or judging in hindsight that they would have been correct/incorrect) in changes to prospective confidence. Affective responses appear to be important for people to learn prospective confidence and to learn to recognize errors in decision making. Greater positive affect to positive feedback was associated with increases in prospective confidence. On the other hand, greater negative affect to negative feedback was associated with more sensitive retrospective confidence judgments (i.e., participants' retrospective confidence judgments differed more for correct and incorrect responses). Drift diffusion modeling revealed that hindsight training (versus foresight-with-feedback) reduced participants' rate of evidence accumulation and their cautiousness in subsequent decisions. Moreover, greater prospective confidence was associated with faster evidence accumulation and more cautious decision making.

CURRICULUM VITAE

NAME OF AUTHOR: Michael C. Silverstein

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
The Ohio State University, Columbus, OH
Muhlenberg College, Allentown, PA

DEGREES AWARDED:

Doctor of Philosophy, Psychology, 2025, University of Oregon
Master of Science, Psychology, 2021, University of Oregon
Bachelor of Arts, Psychology with Honors, Statistics Minor, 2018, Muhlenberg College

AREAS OF SPECIAL INTEREST:

Judgment and Decision Making

GRANTS, AWARDS, AND HONORS:

Knight Campus Undergraduate Scholars Mentor, Knight Campus for Accelerating Scientific Impact, University of Oregon, 2024
Research Pilot Grant, Center for Science Communication Research, University of Oregon, 2020, 2024, 2025
Professional Development Graduate Scholarship, Division of Graduate Studies, University of Oregon, 2023
Gregores Research Award, Department of Psychology, University of Oregon, 2023
Open Access Article Processing Charge Award, University of Oregon Libraries, 2023
First Year Merit Award, Department of Psychology, University of Oregon, 2020
Graduate Student on NSF RAPID Grant, National Science Foundation, 2020
Pilot Grant, Decision Sciences Collaborative, The Ohio State University, 2018
University Fellowship, The Ohio State University, 2018

PUBLICATIONS:

Silverstein, M. C., Bjälkebring, P., Shoots-Reinhard, B., & Peters, E. (2023). The numeric understanding measures: Developing and validating adaptive and nonadaptive numeracy scales. *Judgment and Decision Making*, 18, E19. <https://doi.org/10.1017/jdm.2023.17>

Silverstein, M. C., Lee, C. N., Scherer, L. D., Phommasathit, C., Merrill, A. L., & Peters, E. (2023). Operating on anxiety: Negative affect toward breast cancer and choosing contralateral prophylactic mastectomy. *Medical Decision Making*, *43*(2), 152–163. <https://doi.org/10.1177/0272989X221121134>

Shoots-Reinhard, B., Goodwin, R., Bjälkebring, P., Markowitz, D. M., **Silverstein, M. C.**, & Peters, E. (2021). Ability-related political polarization in the COVID-19 pandemic. *Intelligence*, *88*, 101580. <https://doi.org/10.1016/j.intell.2021.101580>

Markowitz, D. M., Shoots-Reinhard, B., Peters, E., **Silverstein, M. C.**, Goodwin, R., & Bjälkebring, P. (2021). Dehumanization during the COVID-19 pandemic. *Frontiers in Psychology*, *12*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.634543>

ACKNOWLEDGMENTS

First, I wish to express sincere appreciation to Dr. Ellen Peters for her constant support through my graduate training. She is the intellectual and academic giant whose shoulders I stand on. Ellen has served many roles to help me achieve my goals: resource getter, brainstormer, connector, constructive critic, mentor, guide, and friend. Her impact on me as a scientist and a person has been invaluable. I am also grateful to Martin Tusler, whom I always looked forward to chatting with at lab events.

I also want to thank my committee members, Dr. Paul Slovic, Dr. Robert Mauro, and Dr. John Clithero for their contributions to my intellectual development. They have provided helpful feedback and creative suggestions at various stages of this research and my training. Additionally, I want to thank Paul for being one of the first people to welcome me to Eugene. His and Roz's hospitality exemplified the kindness that I would find in Eugene.

I want to express my sincerest gratitude to past and present members of the CAIDe Lab. This lab has been my family-away-from-family for the past 7 years. My research was made possible by the collaborations with and feedback from lab members and the many undergraduate research assistants who have tested and collected my studies. I am especially honored to have had the opportunity to mentor research assistants through their independent research projects.

I want to thank Dr. Hal Arkes for his contributions, as well. The present projects were, in part, inspired by his work on hindsight bias. Further, he was always willing to meet with me to discuss ideas for studies and to share manuscripts and books from his personal collection that helped me further develop my ideas. I'd also like to thank Dr. Jenefer Husman and Dr. Matthew Graham for loaning me their Empatica units.

Lastly, I want to thank my family:

My mother, for her constant reassurance and for teaching me what it is to be a scientist;
My father, for coaching me in all endeavors and pushing me to be my best;
Marlee, for leaving me her footsteps to follow and empowering me to forge my own path;
Steve, for showing me the value of hard work—I'm glad I failed to scare you away;
The Rait family, for accepting me into their family with open arms;
Lindsay, for being my greatest partner and supporter in all aspects of my life. When I was moving to Oregon, I had no idea that I would meet my bashert.

The investigations in this dissertation were supported in part by a pilot grant from the Center for Science Communication Research to Michael Silverstein and by a grant from the National Science Foundation, SES-2017651 (Numeric self-efficacy, objective numeracy, and overconfidence), to Dr. Ellen Peters at the University of Oregon.

DEDICATION

To my parents, who didn't name me 'Doctor,' so I had to earn it.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION AND LITERATURE REVIEW	15
What is Confidence.....	16
Estimation, Placement, and Precision.....	17
Sensitivity and Bias.....	19
Retrospective Confidence versus Prospective Confidence.....	20
The Confidence Cycle.....	22
Prospective Confidence Influences Decision Making.....	23
Decision Processes Inform Retrospective Confidence.....	26
Prospective Confidence Informs Retrospective Confidence.....	28
Retrospective Confidence Informs Prospective Confidence.....	29
Role of Affect in the Confidence Cycle.....	31
Hindsight Bias and Confidence.....	31
Hindsight Bias and The Confidence Cycle.....	35
General Hypotheses.....	37
The Present Dissertation.....	39
II. PILOT STUDIES	42
Pilot Study 1.....	42
Method.....	43
Results and Discussion.....	45
Pilot Study 1 Summary.....	48
Pilot Study 2.....	49
Method.....	50
Results and Discussion.....	53
Pilot Study 2 Summary.....	58
Pilot Study 3.....	59
Method.....	60
Results and Discussion.....	63
Pilot Study 3 Summary.....	68
General Discussion of Pilot Studies.....	69
III. STUDY 1	72
Research Design.....	74
Data Collection.....	80
Data Cleaning.....	80
Data Analysis.....	81
Discussion.....	104
Limitations.....	110
Significance and Contributions to the Field.....	111
IV. STUDY 2.....	114
Research Design.....	114
Data Collection.....	125
Data Cleaning.....	125

Data Analysis	126
Discussion	152
Limitations	158
Significance and Contributions to the Field.....	160
V. GENERAL DISCUSSION	163
Consequences of Hindsight for Subsequent Confidence Judgments.....	166
Functions of Prospective Confidence	169
Affect and Confidence	172
Conclusion	174
REFERENCES CITED.....	176

LIST OF FIGURES

Figure		Page
1.1	The Confidence Cycle	23
2.1	Example of a choice between two expressions	43
2.2	Example of a choice between two expressions with the larger expression indicated.	44
2.3	Participants' accuracy by training condition	46
2.4	Training retrospective confidence by training condition	47
2.5	Prospective confidence by training condition	48
2.6	Predicted marginal effect of prospective confidence on trial-by-trial response times	54
2.7	Predicted marginal effect of prospective confidence on trial-by-trial retrospective confidence.....	55
2.8	Predicted marginal effects of prospective confidence and trial-by-trial accuracy on trial-by-trial retrospective confidence	56
2.9	Predicted marginal effects of prospective confidence and trial-by-trial response times on trial-by-trial retrospective confidence.....	57
2.10	Predicted marginal effects of numeracy and trial-by-trial accuracy on trial-by-trial retrospective confidence	58
2.11	Participants' accuracy by training and difficulty conditions.....	64
2.12	Change in prospective confidence by training and length conditions.....	66
2.13	The interaction of condition and the evaluation subscale of metacognitive awareness predicting training retrospective confidence bias	67
3.1	Average proportion of correct answers indicated during training trials by training condition and training length	84
3.2	Average proportion of correct answers during test trials by training condition and training length.....	85
3.3	Average change in prospective confidence following training by training condition and training length.....	87
3.4	Average overall retrospective confidence judgments for the testing phase by training condition and training length	89
3.5	Mediation model depicting indirect effects of training condition.....	90
3.6	Estimated marginal means of trial-by-trial retrospective confidence by training condition and response time.....	91
3.7	Estimated marginal means of trial-by-trial retrospective confidence by prospective confidence and response time (separate lines represent -1 standard deviation, the average, and +1 standard deviation; A) and by overall test accuracy and trial-level accuracy (B)	93
3.8	Estimated marginal means of prospective confidence bias for the test trials by condition, length, and monitoring	95
3.9	Estimated marginal means of retrospective confidence bias for the test trials by condition, length, and monitoring	96
3.10	Estimated marginal means of response times for the test trials by condition	

	and overall test accuracy	97
3.11	Estimated marginal means of response times for the test trials by prospective confidence	98
3.12	Estimated marginal means of trial-by-trial response times by prospective confidence and trial-level accuracy (A) and by overall test accuracy and trial-level accuracy (B)	100
3.13	Average prospective confidence judgments during a one-week follow-up for a hypothetical test by training condition and training length	102
3.14	Average confidence judgments throughout the study by training condition	103
3.15	The Confidence Cycle	106
4.1	Depiction of a Drift Diffusion Model and its parameters	117
4.2	Depiction of each training trial	120
4.3	Expected marginal means of prospective confidence by test accuracy and training condition	132
4.4	Expected marginal means of retrospective confidence by test accuracy and training condition	133
4.5	Estimated marginal means of affective valence by feedback and training condition	134
4.6	Estimated marginal means of affective valence by feedback and training condition and evaluation tendency	135
4.7	Estimated marginal means of affective arousal by feedback and training condition	136
4.8	Estimated marginal means of change in prospective confidence by affect valence to positive feedback and training accuracy	138
4.9	Parallel mediation model depicting indirect effects of training condition on prospective confidence bias	139
4.10	Serial mediation model depicting indirect effects of the training condition on retrospective confidence for correct and incorrect answers	141
4.11	Serial mediation model depicting indirect effects of the training condition on retrospective confidence bias and sensitivity	143
4.12	Estimates of marginal means of trial-by-trial retrospective confidence by trial accuracy and affective valence to negative feedback	145
4.13	Estimates marginal means of average retrospective confidence by overall test accuracy and affective valence to negative feedback	145
4.14	Depiction of estimated DDM parameters when prospective confidence is 1 standard deviation above average (A), average (B), and 1 standard deviation below average (C)	151

LIST OF TABLES

Table		Page
2.1	Descriptive statistics of training retrospective confidence by training condition	53
2.2	Descriptive statistics of prospective confidence by training condition	53
2.3	Sample sizes (n) by condition	61
3.1	Descriptive statistics of Study 1 variables and demographics	82
3.2	Study 1 Pearson correlations of between-subjects variables and indices	83
4.1	Descriptive statistics of Study 2 variables and demographics	127
4.2	Study 2 Pearson correlations of between-subjects variables and indices	128
4.3	Significant indirect effects of training condition on retrospective confidence for correct and incorrect answers	141
4.4	Significant indirect effects of training condition on retrospective confidence bias and sensitivity	143
4.5	Fitted DDM parameter estimates by training condition	149
4.6	Fitted DDM parameter estimates by prospective confidence	150

I: INTRODUCTION AND LITERATURE REVIEW

When I was younger, I struggled with vocabulary and spelling in school. My parents encouraged me to study and practice for my vocabulary tests with flashcards (e.g., an index card with a question on the front and the correct answer on the back). Flashcards allow you to quiz yourself in the hopes of identifying areas of mastery and areas that need further review. Of course, flashcards work best if you use them as intended:

- 1) Look at the question, 2) answer the question, 3) check the answer provided against the true answer, and 4) repeat steps 1-3 for all cards.

However, I was impatient and easily tempted to look at the answers, so I would cheat a bit.

Instead of the intended use, I would:

- 1) Look at the question, 2) look at the true answer, 3) judge if I would have known the correct answer, and 4) repeat steps 1-3 for all/most/some cards.

This latter strategy required that I judge whether I would have been correct in *hindsight*, after knowing the true answer. Often, I would judge that I would have been correct when I likely would not have, and, after going through the pile of flashcards, I would feel confident about my vocabulary and spelling for the test. However, I would, ultimately, do poorly on the test. This outcome was especially crushing, since I typically felt that I had done pretty well immediately after the test. Apparently, studying in a way that relied on judgments in hindsight opened me up to bias and made me overconfident. Each time that I looked at the answer before attempting to answer first, I overestimated the extent to which I knew the correct answer. With each additional

card, I felt more and more confident in my vocabulary and spelling, and that I was ready for the test.

The tendency to overestimate the extent you would have known/predicted something after seeing the outcome or correct answer is referred to as *hindsight bias* (Fischhoff, 1975). Studying in a way that affords a hindsight bias (e.g., looking at the answers before attempting the question) may cause people to retrospectively overestimate their knowledge, in other words, to become overconfident (Fischhoff, 1975; Winman et al., 1998). This lack of calibration itself can be problematic. However, the consequences of hindsight bias—especially repeated experiences of hindsight bias—for future decision making have been speculated but little explored. For example, it is possible that hindsight bias, through efficacy beliefs (i.e., beliefs about one’s ability; e.g., “I am good at math”) and confidence (i.e., judgments about one’s performance on a specific task; e.g., “I believe I will answer that math problem correctly”), might influence how people decide subsequently and bias *prospective confidence* as well as *retrospective confidence* for performance in subsequent similar tasks. Hindsight bias and its consequences for prospective and retrospective confidence are the primary focus of this dissertation. The secondary focus of this dissertation is to explore a possible Confidence Cycle (see *The Confidence Cycle* later in this chapter)—particularly the effects of prospective confidence on decision making and how we learn prospective confidence through affect.

What is Confidence?

When people talk about confidence, they might refer to a judgment made prior to a decision (i.e., prospective) or after the fact (i.e., retrospective). However, these prospective and retrospective

judgments appear to be relatively distinct, differing in sources (i.e., the cognitive and experiential bases of the judgment; Fleming et al., 2016) and associated brain regions (Fleming & Dolan, 2012). Here, **decision confidence** is defined as *a decision-specific metacognitive likelihood judgment of decision-goal attainment (e.g., accuracy) made prior to a decision (i.e., prospective) or following a decision (often in the absence of objective feedback; i.e., retrospective). These judgments are predicated on efficacy beliefs, formed based on learned task difficulty cues, and informed by the decision-making experience (e.g., fluency, decision time)*. Much of the cognitive and decision-making literature on confidence has focused largely on retrospective confidence in decisions, with a strong emphasis in decision research on overconfidence (e.g., Moore & Healy, 2008). Social psychology instead has had a focus on beliefs about one's ability prior to behavior (e.g., efficacy beliefs; Bandura, 1977). Memory researchers have studied how people know what they know prospectively in the area of metamemory (Koriat, 2007). Recently, cognitive psychology in perceptual decision making has begun to investigate such prospective beliefs (e.g., Boldt et al., 2019; Fleming et al., 2016). The present review pulls from these various bodies of literature to discuss the interrelations of confidence with decision making and the potential consequences of hindsight bias for subsequent confidence judgments and decisions.

Estimation, Placement, and Precision. Broadly, confidence judgments can be separated into three varieties: placement, precision, and estimation (see Moore & Healy, 2008). While the focus of the present dissertation is *estimation* judgments, we now briefly discuss placement and precision. Placement confidence judgments are judgments of performance or ability relative to other people. An example of placement would be if people are asked to estimate their rank order or percentile rank on an exam performance relative to their classmates. Pervasive overplacement

(i.e., people rate themselves as better relative to others than they are) results in the “better-than-average effect”—in which the majority of people think they are above the 50th percentile (see Alicke & Govorun, 2005). Precision judgments can be assessed for beliefs that can be expressed quantitatively (e.g., How long is the Nile River) by requesting a confidence interval (typically a 90% confidence interval; Moore & Healy, 2008). More confident people should provide narrower confidence intervals than those who are less confident. Typically, precision is assessed in the context of knowledge, and participants are asked to provide a range of possible values that they are 90% confident contains the correct value. People are typically over-precise, meaning that their 90% confidence interval contains the correct value less than would be expected for an accurate 90% confidence interval—some studies suggest only 50% of the time (Soll & Klayman, 2004).

The focus of the present dissertation is on estimation confidence judgments. These are *judgments of one’s actual ability, performance, level of control, or chance of success* (Moore & Healy, 2008). Estimation can be elicited for past decisions (retrospectively) or future tasks (prospectively; Fleming et al., 2016). For retrospective judgments, people are often asked to report their confidence about their response or responses on a scale from 0%-100% (alternatively from 50%-100%) for a single decision they just made. Alternatively, they can be asked to report their confidence about their answer on an ordinal Likert-type scale (e.g., 1- Not at all Confident to 6- Very Confident). Sometimes, people are asked to report their confidence and answer using a single semantic differentiation scale (e.g., 1- Certainly No to 6- Certainly Yes). After a series of decisions, people could be asked to estimate how many questions they think they got correct out of a series of decisions. For prospective judgments, many of these same types of self-report measures apply. People could be asked how many questions they expect to get correct or to

estimate the likelihood that they will be correct for a specific decision. Alternatively, people could be asked about their ability or their efficacy beliefs. For example, people could be asked to report how good or bad they are at a task. Prospective estimation judgments about ability could be thought of as similar to self-efficacy. According to Bandura (1977), self-efficacy is domain-specific confidence in one's ability to be successful.

Sensitivity and Bias. When discussing the correspondence of confidence judgments to accuracy, it is important to distinguish between two aspects of confidence calibration: sensitivity and bias (Fleming & Lau, 2014; Koriat, 2007). Confidence sensitivity represents the ability of a decision maker to discriminate between good and bad decisions in their confidence judgments and is related to the confidence-accuracy correlation. Whereas bias is related to over- and under-confidence—individual differences in task confidence—and is often referred to as accuracy. Sensitivity is the difference between a person's average confidence when they are correct and average confidence when they are incorrect. Accuracy, on the other hand, is the difference between a person's average confidence and their average accuracy.

$$Sensitivity = \bar{C}_{O=1} - \bar{C}_{O=0}$$

$$Bias = \bar{C} - \bar{O}$$

For example, if a person completes ten math problems and reports their confidence following each question (on a 0 [certain incorrect]-100 [certain correct] scale), we can assess the sensitivity and accuracy of their confidence judgments. Imagine that they answered five of the ten correctly. If they are highly confident when they are correct—on average indicating a 95 for their confidence—and not at all confident when incorrect—on average indicating a 45 for their confidence—, their confidence is sensitive to their accuracy. However, if they report the same

confidence level regardless of their accuracy (say they indicate 50 for their confidence for every problem), their confidence is not at all sensitive to their accuracy. However, in this latter case, the person's confidence judgment would be accurate—their average confidence would equal their average correctness. In the former case, however, the person would be overconfident because their average confidence (70) would be greater than their average accuracy (50%).

Generally, people's retrospective confidence judgments are somewhat reflective of their accuracy in that confidence ratings correlate positively with accuracy, even in the absence of feedback, with and without controlling for difficulty (Ariely et al., 2000). That is, people tend to be more confident when they are correct and less confident when they are incorrect—they demonstrate sensitivity in their confidence judgments. However, people may lack accuracy—often demonstrating overconfidence and underconfidence in different situations. This decomposition of confidence into sensitivity and accuracy also applies to prospective confidence judgments but has rarely been performed. However, Fleming and colleagues (2016) found that retrospective confidence was more sensitive and accurate than prospective confidence—likely because retrospective confidence can take into account the specific decision-making experience.

Retrospective Confidence versus Prospective Confidence

Aside from the timing of the confidence judgment relative to a decision, what else distinguishes prospective and retrospective confidence? There is some evidence that prospective and retrospective confidences are neurologically distinct (Chua et al., 2009; Fleming & Dolan, 2012). Research in metamemory, for example, has found that right ventromedial prefrontal cortex damage was associated with a decreased accuracy in feeling-of-knowing judgments (i.e., prospective confidence of recall ability) but did not affect retrospective confidence accuracy

(Schnyer et al., 2005). On the other hand, lateral frontal lesions were associated with impairment of retrospective confidence, but not prospective confidence (Pannu et al., 2005). Functional MRI studies have suggested that prospective confidence is associated with activity in the medial prefrontal cortex (PFC; Modirrousta & Fellows, 2008; Schnyer et al., 2004), while retrospective confidence is associated with activity in the rostromedial PFC (Fleck et al., 2006; Fleming et al., 2012; Yokoyama et al., 2010). Interestingly, Euston and colleagues (2012) suggest that the medial PFC learns and stores associations between events, context, locations, and corresponding affective responses. The medial PFC has also been indicated in the learning of affective responses to inform decisions (Bechara & Damasio, 2005; see *Role of Affect in the Confidence Cycle* below).

Prospective and retrospective confidence also appear to be somewhat distinct in terms of their sources and functions. Prospective confidence is thought to draw largely from past experiences to inform expectations (Bandura, 1977; Feather, 1966; Fleming et al., 2016), guide the allocation of effort and help seeking (Aguilar-Lleyda et al., 2020; Arkes et al., 1986; Bruine de Bruin et al., 2023; de Gardelle & Mamassian, 2014; Parker & Stone, 2014; Pescetelli et al., 2021; Sieck & Arkes, 2005), and motivate/demotivate deliberation (see *Prospective Confidence Influences Decision Making* below). Retrospective confidence, in contrast, is thought to draw largely from the specific, just-completed decision-making experience (Koriat et al., 1981; Koriat & Adiv, 2012; Koriat & Levy-Sadot, 1999) to inform decision quality (Carlebach & Yeung, 2020; Yeung & Summerfield, 2012).

Prospective confidence has been little considered in the study of judgment and decision making. Instead, decision research has focused primarily on judgments of confidence about past decisions with an emphasis on metacognitive bias (e.g., overconfidence) and improving calibration. The

study of prospective confidence has largely been left to social psychology (e.g., self-efficacy) and metamemory research (e.g., feelings of knowing and judgments of learning). However, prospective and retrospective judgments appear to be separate but related constructs. It is likely that prospective confidence directly informs retrospective judgments (Ehrlinger & Dunning, 2003; Van Marcke et al., 2024) and those retrospective judgments, in turn, inform prospective confidence—possibly through efficacy beliefs (Bandura, 1977; Feather, 1966; Fleming et al., 2016). Importantly, prospective confidence has been linked to some extent to decision processes (e.g., Boldt et al., 2019; Desender et al., 2019) and outcomes (Peters, Tompkins, et al., 2019; Peters & Shoots-Reinhard, 2022). This dissertation seeks to investigate the relations between prospective confidence, decision-making, and retrospective confidence. Moreover, the present dissertation will investigate how hindsight bias affects subsequent decision-making and confidence calibration through these relations.

The Confidence Cycle

We propose that there is a Confidence Cycle (see Figure 1.1). In the following sections, we review literature that suggests that 1) prospective confidence influences decision making (path A), 2) decision making influences retrospective confidence (path B), 3) prospective confidence informs retrospective confidence (path C), and 4) retrospective confidence informs prospective confidence (path D). We will also discuss the potential role of affect in the Confidence Cycle. In the following chapters, we will extend findings about the role of prospective confidence in decision making beyond perceptual decisions—to a novel numeric decision task.

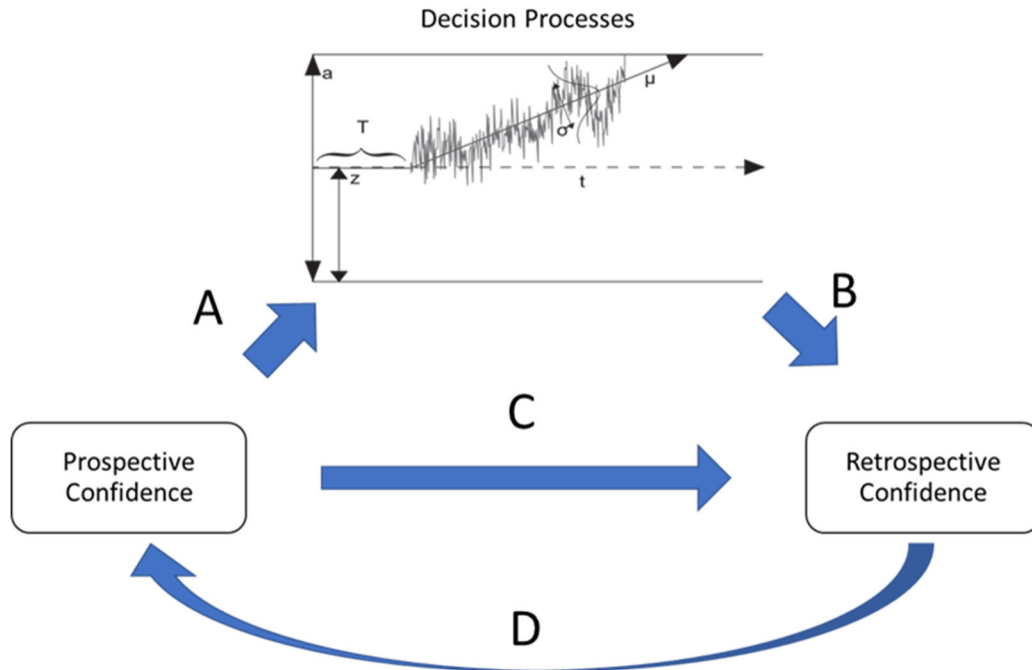


Figure 1.1. *The Confidence Cycle. Decision Processes as depicted in line with the Drift Diffusion Model (DDM; Ratcliff, 1978). The DDM suggests that a binary choice can be represented as a random evidence accumulation toward decision thresholds and consists of multiple parameters: 1. Drift rate (μ): The rate at which evidence accumulates for one option relative to another. 2. Decision threshold (a): The amount of evidence required to decide. This parameter is sometimes referred to as decision cautiousness. 3. Starting point (z): The point from which the evidence accumulation process begins relative to the decision thresholds. 4. Non-decision time (T): The time required for stimulus encoding, motor preparation, and other non-decision-related processes. 5. Variability of evidence (σ): The variability of the evidence for each option. A higher variability of evidence makes it harder to distinguish the signal from the noise.*

Prospective Confidence Influences Decision Making. Critical to the outcome of decision making, prospective confidence modulates effort and deliberation. Research in the area of metamemory suggests that feelings of knowing and judgments of learning dictate the allocation of resources in memory retrieval and the selection of learning strategies (Koriat, 2007). For example, prospective confidence predicts the length of time studying items for and answering items on a recall task (Son & Metcalfe, 2000), time spent searching for a memory (Singer & Tiede, 2008), and the allocation of studying over time (Son, 2004). Additionally, feelings of

knowing motivate strategies for answering questions—recalling them from memory or deriving them through logic (Reder, 1987; Reder & Ritter, 1992).

Inspired by the metamemory literature, disfluency was suggested as a trigger of more effortful, deliberative processing since experienced disfluency can cue task difficulty (Alter et al., 2007; Alter & Oppenheimer, 2009). Zhao and colleagues (2022) consistently found that disfluent fonts increased cautiousness (i.e., threshold separation) modeled using Drift Diffusion Modeling (DDM), suggesting that disfluent processing triggered participants to be more cautious and deliberative. Similarly, Thompson and colleagues (2011) suggested that, in a default-interventionalist framework of decision making, “feelings of rightness” accompanying an intuitive response could motivate the intervention of more deliberative processing. Both theories suggest that reducing prospective confidence might increase deliberation. Consistently, low confidence in a decision has been shown to increase deliberation in a subsequent decision (Desender et al., 2019), and reducing prospective confidence has been shown to increase effort and deliberation (Stone, 1994). Moreover, decision makers become more cautious (i.e., expanding thresholds in DDM) following a perceived error (Dutilh et al., 2012). Vancouver and colleagues (2002) found that repetition of a task induced overconfidence and increased logical errors, potentially because of less caution in decision making. All these findings suggest that *greater prospective confidence decreases deliberation*.

However, other theories have suggested the opposite is true (Boldt et al., 2019; van den Berg et al., 2016; Zacharopoulos et al., 2014): that increasing *prospective confidence increases deliberation*. Boldt and colleagues (2019) used learned difficulty cues to manipulate prospective confidence for a perceptual decision task. They found that trials cued to be easy caused participants to invest more effort (i.e., they expanded decision thresholds according to DDM) and

prepare more (i.e., they showed greater preparatory EEG activity). Similarly, van den Berg and colleagues (2016) found that participants adjusted their deliberation on a second decision based on their confidence in the first, such that higher confidence was associated with more cautious decision making. One explanation for more effort on easy trials is that participants may allocate their effort to trials they expect to be able to succeed at to conserve energy (Shenhav et al., 2021). In this way, prospective confidence may act as a self-fulfilling prophecy (Zacharopoulos et al., 2014). Research on numeric self-efficacy suggests that more confident people persist more on numeric tasks, driving numeric deliberations and calculations (Peters & Shoots-Reinhard, 2022).

There appear to be opposite findings regarding the relation between prospective confidence and deliberation. These findings beg the question: when does confidence motivate effortful deliberation and when does it demotivate? A classical theory of achievement motivation posits that motivation for a given task is a function of the value of the outcome and the perceived likelihood of task achievement (Atkinson, 1957). Of particular importance is that the function between motivation and perceived likelihood of achievement is not linear, but rather parabolic, with the greatest motivation occurring when the task difficulty is equal to one's perceived ability (i.e., a 50-50 chance of success). When motivated by achievement goals, people may deliberate more when their perceived uncertainty of success is highest and deliberate less when they view the task as beneath them or over their heads. The achievement motivation model suggests that when perceived task difficulty is high, highly confident people may be motivated to deliberate. Such deliberation may be in vain, though, if the task is too difficult for them. This is the common understanding of the negative influence of being overconfident: colloquially, one might "bite off more than one can chew." However, little research has evaluated situations in which highly

confident people are overconfident, such that they feel a difficult task is beneath them. This may occur when a task appears easy but is deceptively difficult. The achievement motivation model would suggest that, in this situation, such people may be demotivated to deliberate. We further suggest that people will rely more on intuition and may make worse decisions when deliberation is demotivated. Hence, if there is more to a task than meets the eye, a person might believe it is easy (when it is not) and thus abstain from deliberation. This idea that overconfidence leads to carelessness is not new; indeed, Stone (1994) found that when negative expectations were induced to counter overconfidence, participants increased effort, deliberation, and performance in a decision task. It is also possible that focusing on the process of goal pursuit (e.g., trying to enjoy the task itself) rather than the goal itself (e.g., viewing the task as just a means to an end) may cause uncertainty in outcomes to be motivating (Shen et al., 2014). In Studies 1 and 2, we will explore a potential non-linear relation between prospective confidence and deliberation.

Decision Processes Inform Retrospective Confidence. Some research suggests that retrospective confidence is *information-based*—formed based on analytic processes used to estimate a likelihood of accuracy (Koriat & Levy-Sadot, 1999). That is, people infer their likelihood of correctness based on the balance of evidence recruited in support of the options (following the decision-making process), their metacognitive theories (e.g., I make worse decisions when rushed or tired), and/or recollection of similar tasks and prospective confidence (more on prospective confidence in *Prospective Confidence Informs Retrospective Confidence* below). For example, Koriat, Lichtenstein, and Fischhoff (1981) suggested that following a decision, people review the evidence used to decide to form a judgment of correctness. However, they thought that this review process was biased by attempts to justify the decision made; people focus on evidence consistent with their decision. Similarly, Kruger and Dunning (1999) suggested that

people rely on the same ability to inform confidence judgments for a given task and to complete the task. Additionally, people report being less confident in their answers when they believe that the evidence used to decide was unreliable (Boldt et al., 2017). People greater in metacognitive awareness tend to attend to whether they are effectively achieving and pursuing their goals more and have confidence judgments more accurate to their actual performance (Stankov & Lee, 2008).

There is also evidence of retrospective confidence being *experience-based*—inferred based on heuristic cues derived from the decision-making experience (Koriat & Levy-Sadot, 1999). These cues are thought to automatically give rise to feelings of confidence. Commonly cited heuristic cues for decision accuracy involve fluency (see Alter & Oppenheimer, 2009). Greater fluency is associated with greater retrospective confidence judgments regardless of whether the source of fluency is perceptual (Alter et al., 2007), based on encoding processes (Castel et al., 2007; Koriat & Ma'ayan, 2005), or based on retrieval processes (Kelley & Lindsay, 1993). For example, Alter and colleagues (2007) found that people were less confident in their answers to logic problems when the instructions and questions were presented in a difficult-to-read font. They interpreted these findings as evidence that disfluency in perceptual processing acted as a difficulty cue, reducing retrospective confidence judgments. How long a decision takes can also heuristically inform retrospective confidence. When people are allowed to respond at their own pace, they tend to report greater confidence in fast responses (Baranski & Petrusic, 1998). However, when decision time or evidence presentation is manipulated (e.g., in an interrogation paradigm or by manipulating speed v. accuracy goals), subjects tend to be more confident in decisions that take longer (Vickers et al., 1985). This pattern of results suggests that greater accumulated evidence results in greater confidence; however, time can be a heuristic cue for difficulty and thus

confidence. When evidence is accumulated slowly and the decision takes longer than expected, people are less confident in their decision.

Prospective Confidence Informs Retrospective Confidence. Moore and Healy (2008)

conceptualized retrospective confidence judgments as a process like Bayesian belief updating in which people base their judgments on a metacognitive theory about the probability distribution of performance on a task and update based on gut feelings. In this way, people's retrospective confidence judgments are based on imperfect knowledge and regress toward their subjective beliefs about the probability distribution of performance. Crucial to this Bayesian theory of confidence is that people hold a metacognitive theory that their gut feelings convey an imperfect indication of performance and that the true performance must exist between the feeling and their prior beliefs. These prior beliefs about performance are somewhat like prospective confidence judgments.

Indeed, prospective confidence and self-efficacy appear to be a source of information for retrospective confidence (Bandura, 1977; Dunning et al., 2003). Ehrlinger and Dunning (2003) found that people's retrospective confidence judgments varied based on what they believed the test was measuring—aligning with how confident they felt in their own abilities in that domain. When people believed the test assessed abstract reasoning (an ability on which participants rated themselves highly), they believed they had done better than participants who believed the same test assessed computer programming (an ability on which participants did not rate themselves highly). In reality, groups did not differ in their performance regardless of their belief of what the test assessed, suggesting the difference in confidence was due to the difference in their prospective confidence. Bolt and colleagues (2019) found in a perceptual decision task that learned difficulty cues predicted retrospective confidence even when the cue-difficulty pairs

were changed unannounced. These results suggest that expectations learned from experience inform retrospective confidence judgments. More recently, Van Marcke and colleagues (2024) manipulated prospective confidence through false feedback during a training phase of a study and demonstrated changes in retrospective confidence bias. False feedback during training indicating that the participants were performing worse or better than they truly were caused underconfident and overconfident retrospective judgments of performance on subsequent perceptual decisions. Thus, prospective confidence may serve as a pseudo-Bayesian prior from which we update retrospective judgments using the decision-making experience. The effects of prospective confidence on retrospective confidence judgments at the decision level have not yet been shown for non-perceptual decisions.

Retrospective Confidence Informs Prospective Confidence. Bandura (1977) identified four main factors that influence self-efficacy: 1) personal experience of success and failure, 2) vicarious experience (i.e., observing others' successes or failures), 3) social persuasion (e.g., encouragement from others), and 4) physiological factors (e.g., fatigue). The strongest factors identified by Bandura are personal and vicarious experiences. Prior experiences of success and failure inform expectations of future outcomes (Bandura, 1977; Feather, 1966; Fleming et al., 2016). In the absence of veridical feedback, people appear to base prospective confidence judgments on their past retrospective confidence (Wittmann et al., 2016; Rouault et al., 2019, 2023). People can use retrospective confidence to determine abstract features of their environment (e.g., reward rate, difficulty; Wittmann et al., 2016). For example, people can learn difficulty cues for tasks without feedback (Boldt et al., 2019). They can also use retrospective confidence to inform beliefs about their efficacy (Rouault et al., 2019, 2023). However, people do not appear to update their efficacy beliefs in an unbiased way. Relative to what would be

expected from Bayesian updating, people appear to overweight recent experiences (Fleming et al., 2016), are conservative in their updates (Möbius et al., 2022), and demonstrate asymmetric updating—they overweight positive experiences relative to negative experiences. Moreover, experience itself, regardless of perceived success and failures, may inflate prospective confidence (Einhorn & Hogarth, 1978). For example, Dawson and colleagues (1993) found that physicians with more experience performing right-heart catheterizations were more confident in their ability to estimate the information provided by the procedure, despite never making such estimates and being no more accurate than less experienced physicians. One potential reason for this inflation is that repeated hindsight bias could have provided false positives. Supporting this explanation, Chance and colleagues (2011) found that giving participants answers to a test as they took it increased their confidence for another test for which they wouldn't be given the answers. Even when given veridical feedback about their accuracy, decision makers might become more confident with experience, especially when decisions are based on multiple probabilistic cues (Brehmer, 1980). Vicarious experiences can also inflate confidence (Bandura, 1977). For example, Kardas and O'Brien (2018) found that watching others perform or explain how to perform a task (e.g., throwing darts, juggling) increased confidence without increasing ability.

In the present dissertation, we will attempt to interfere with participants' ability to learn from retrospective confidence during training by affording them hindsight bias repeatedly. We suspect that repeated experiences of hindsight bias will increase prospective confidence and, ultimately, increase retrospective confidence (driving retrospective confidence bias) and affect decision processes for subsequent decisions.

Role of Affect in the Confidence Cycle. The valence of affect has been shown to inform retrospective confidence, with positive moods increasing confidence (Ifcher & Zarghamee, 2014; Koellinger & Treffers, 2015) and negative moods decreasing confidence (Culot et al., 2021). Positive moods have also been associated with greater self-efficacy (see Lyubomirsky et al., 2005) and increased prospective confidence for laboratory tasks (Wright & Mischel, 1982). However, these studies have been limited to investigating incidental affect. In Study 2, we will investigate the role of anticipatory affect during training on how people learn confidence. To the extent that prospective confidence reflects prior successes and failures, affect may play a crucial role in the indexing of experience, storage, and retrieval of previous experiences. Lebreton and colleagues (2015) found that the ventromedial prefrontal cortex (PFC) automatically integrates retrospective confidence in encoding values of choices, and multiple functional MRI studies have suggested that prospective confidence is associated with activity in the medial prefrontal cortex (Modirrousta & Fellows, 2008; Schnyer et al., 2004). Given the involvement of the medial PFC in prospective confidence, affective responses following perceived failures and successes may be critical for learning confidence in a task (Bechara & Damasio, 2005; Damasio, 1994). Additionally, experienced emotions at the time of forming prospective confidence judgments may influence what experiences are recalled to inform the confidence judgment (Bower, 1981).

Hindsight Bias and Confidence

In the early 90's, many physicians who regularly conducted right-heart catheterizations felt that it was unnecessary to continue performing right-heart catheterizations on critically ill patients. They felt that, with all their experience, they could accurately estimate the results by observing the patient and through less invasive means, making the procedure more harmful than it was worth for critically ill patients. Dawson and colleagues (1993) investigated whether the

physicians' confidence was warranted and found that more experienced physicians felt more confident in their predictions of right-heart catheterization results. However, they showed that retrospective confidence and experience, while positively related to each other, were both unrelated to accuracy in the predictions. The authors speculated that the effect was due to the knew-it-all-along effect of hindsight and getting outcome information without having predicted the results (in practice, physicians do not state predictions about results before testing). The authors suggested that repeated experiences of hindsight bias had disassociated confidence and accuracy. A possible mechanism would be that physicians, without having made predictions, felt that their prior estimates would have been accurate if they had been made. The outcome information was therefore spuriously positive and increased physicians' confidence in their ability to predict outcomes without any real indication of their ability to predict the outcome. This example, documented by Dawson and colleagues (1993), illuminates the potential pitfalls of repeated hindsight experiences for prospective confidence and retrospective confidence. However, physicians' prospective confidence in predicting these results was not assessed (although anecdotally, prospective confidence was generally high), and the role of hindsight bias was speculative. Nonetheless, this case demonstrates the potential consequences of repeated hindsight experiences and suggests the need for a more longitudinal study of hindsight and confidence. The causal relation between repeated hindsight experiences with prospective and retrospective confidence is the primary focus of this dissertation. Further, the relationship between the amount of experience in hindsight and confidence will also be investigated.

To our knowledge, the assumption that hindsight bias increases confidence for subsequent tasks is common but has not been directly investigated. For example, Hawkins and Hastie (1990, p. 315) assert that hindsight effects "undoubtedly contribute to overconfidence." They reasoned that

if people feel that they would have known an answer or outcome, they would fail to learn from uncertain outcomes and become more confident in their ability and knowledge. Consistently, Bukszar and Connolly (1988) suggested that hindsight bias was a hindrance to learning. In a study of MBA students predicting first-year investment outcomes, some students were shown the outcome but asked to ignore it when forming their own predictions. MBA students failed to disregard the outcome information but nonetheless judged their decision process as adequate, seeing no need to adjust their approach. However, Bukszar and Connolly (1988) did not assess students' confidence in their ability to predict investment returns in the future, nor did they assess retrospective confidence for subsequent, related decisions. It is important to note that outcome information does not solely bias people's judgment of the predictability of outcomes. Hoch and Loewenstein (1989) found that feedback, despite increasing people's confidence that they would have been correct in hindsight, also facilitated people's learning to discriminate between easy and difficult problems.

There is causal evidence that hindsight bias increases confidence in previously given answers (Fischhoff, 1975; Winman et al., 1998) and may increase efficacy beliefs when predictions are confirmed (Bradfield & Wells, 2005). Bradfield and Wells (2005) found that after knowing an answer, participants whose predictions had been confirmed distorted their reports of their certainty about the answer, the quality of their evidence, and the length of time it took them to decide. The same participants also reported being better at related tasks. These findings suggest that people may distort their beliefs about their abilities based on hindsight memory distortions that lead them to believe that they knew it all along and, ultimately, may increase their prospective confidence. However, this study involved participants making a prediction and then being given feedback. It is also possible that the increase in efficacy beliefs was not due to the

hindsight bias but rather to the informative positive feedback (Hoch & Loewenstein, 1989). In contrast, we will investigate hindsight in experiences for which people do not make predictions before learning the true answer. Another gap in the hindsight and confidence literature concerns how durable any confidence increases are. In Study 1, we will investigate the durability of hindsight effects on confidence by eliciting prospective confidence judgments from participants in a one-week follow-up session.

There is no dearth of research suggesting that confidence and other metacognitions are important factors in whether people experience hindsight bias. Winman and colleagues (1998) found that hindsight bias is stronger in tasks for which people tend to be overconfident—that is, difficult ones. Surprise appears to counteract hindsight bias (Hoch & Loewenstein, 1989). When the correct answer is surprising to people, they are less likely to experience hindsight bias.

Processing fluency of the correct answer may also contribute to hindsight bias (Werth & Strack, 2003). When people process the correct answer more fluently, they may erroneously feel it is more familiar and feel that they knew it all along. The influence of fluency suggests that there may be a difference in hindsight bias between people who make a prediction and then receive outcome information and those who immediately receive outcome information without a prediction.

Confidence and other metacognitions have been shown to be important determinants of when hindsight bias occurs (Hoch & Loewenstein, 1989; Werth & Strack, 2003; Winman et al., 1998). When people make a prediction that is confirmed by feedback, people also experience hindsight bias, which may increase efficacy beliefs (Bradfield & Wells, 2005). However, hindsight in the absence of any predictions has only been speculated to increase prospective confidence and retrospective confidence for subsequent similar tasks (e.g., Dawson et al., 1993). Additionally,

the cumulative effects of hindsight bias have also only been speculated. It is possible that when people see an outcome, without having predicted it ahead of time, they will experience hindsight and feel good as if they had been successful. If successes increase self-efficacy and prospective confidence (Bandura, 1977), then their confidence, both prospective and retrospective, will increase without opportunities to improve accuracy. Study 1 seeks to expand on the hindsight and confidence literature in three ways. The first is to investigate if seeing correct answers without having made a prediction increases prospective confidence and retrospective confidence for subsequent decisions. Thus, we test the explanation of physicians' overconfidence in their ability to predict right-heart catheterization outcomes proposed by Dawson and colleagues (1993). The second is to investigate if a greater amount of experience in hindsight further inflates confidence. This research aim is based on the finding that experience was associated with confidence, despite neither being associated with accuracy (Dawson et al., 1993). Thirdly, we will investigate if repeated hindsight experience disassociates retrospective confidence and accuracy—reducing confidence sensitivity. We are also interested in how repeated hindsight experiences affect subsequent decision making.

Hindsight Bias and The Confidence Cycle

In the following chapters, we manipulate whether people experience hindsight bias to examine paths A, C, and D of the proposed Confidence Cycle (see Figure 1.1). We will attempt to manipulate prospective confidence through repeated hindsight bias. When people experience hindsight bias, they often report greater confidence in how they would have performed (Fischhoff, 1975, 2007; Hoch & Loewenstein, 1989; Winman et al., 1998)—a phantom retrospective confidence. We expect this phantom retrospective confidence to act similarly to retrospective confidence experienced following a decision in foresight (we investigate this

mechanism in the pilot studies but not in Studies 1 and 2). Consistently, we expect hindsight bias to increase prospective confidence, particularly when there are repeated hindsight experiences.

By manipulating prospective confidence, we can investigate the effects of prospective confidence on decision processes (path A) and retrospective confidence judgments (path C).

As part of the present dissertation, we are also interested in *how* people learn confidence from experience (path D). Functional MRI studies have suggested that prospective confidence is associated with activity in the medial prefrontal cortex (PFC; Modirrousta & Fellows, 2008; Schnyer et al., 2004)—a region that Euston and colleagues (2012) suggest learns and stores associations between events, context, locations, and corresponding emotional responses. Because activity in the medial PFC is related to emotional associations, emotional responses may mediate the effect of performance feedback on efficacy beliefs and prospective confidence. Emotional responses may be necessary for people to learn and update their confidence. Just as people use affect to inform risk and benefit perceptions (Finucane et al., 2000; Peters et al., 2006), affect may inform confidence judgments, particularly prospective judgments. We propose a hypothesis similar to the somatic marker hypothesis (see Bechara & Damasio, 2005; Damasio et al., 1996): As people gain experience in a task, they have affective responses to the task based on their performance or perceived performance (as informed by objective feedback and confidence). Over time, affective responses are indexed and associated with the task. When a person faces a similar task, they experience anticipatory affect that informs their prospective confidence and guides their interaction with the task. Supporting the idea that affect may be involved in learning confidence and informing prospective confidence is that physiological responses associated with affective arousal (i.e., electrodermal activity [EDA]) tend to fluctuate when students see a difficult problem and receive feedback (Fritz et al., 2014; Horvers et al., 2021; Malmberg et al.,

2019; Wampfler et al., 2019). Interestingly, hindsight might dampen affective responses compared to foresight by mitigating surprise (Mellers et al., 2014). Thus, by affording participants hindsight bias, we might be able to investigate the role of emotions in learning confidence. We will explore the role of affect in learning confidence from experience in Study 2.

Very little research has investigated how confidence is learned over time. Sanchez and Dunning (2018) found that a little experience quickly increased confidence for a novel task (a “beginner’s bubble”) and that this surge in confidence was accompanied by incorrect lay theories of how to best approach the task. In later trials, feedback often disconfirmed and refined these theories, and confidence leveled off while ability improved. When participants are given feedback, they still quickly gain confidence, however, that confidence is tempered by feedback in later trials, suggesting that the amount of experience is important for learning about confidence with feedback. We suspect that when outcome information is provided without a prediction, hindsight bias will prevent later outcomes from attenuating the increase in confidence, since participants will feel that they knew the answer all along.

General Hypotheses

Overconfidence is sometimes thought to be a consequence of cumulative hindsight-bias effects (e.g., Dawson et al., 1993). However, their objective causal relation has gone largely unstudied, particularly when predictions are not made before feedback. We hypothesized that when people see an outcome, without having predicted it ahead of time, they will experience hindsight and feel as if they had been successful; this is the prototypical hindsight hypothesis. If successes increase confidence (Bandura, 1997), then their confidence will increase without opportunities to improve accuracy. The association between more experience and greater confidence (e.g.,

physicians with more years of experience; Dawson et al., 1993) suggests—and we hypothesize—that experiencing hindsight more often will cause even greater confidence.

General Hypothesis 1 and 2 (GH1 and GH2): We propose that having repeated hindsight experiences vs experiences in foresight without feedback will cause, on a subsequent, related task, increased efficacy beliefs and prospective confidence (GH1) and, in turn, greater retrospective confidence (GH2).

General Hypothesis 3 and 4 (GH3 and GH4): We propose that having more vs fewer hindsight experiences will cause, on the subsequent related task, increased efficacy beliefs and prospective confidence (GH3) and, in turn, greater retrospective confidence (GH4).

General Hypothesis 5 (GH5): We further propose that increases in prospective confidence will motivate/demotivate deliberation on subsequent decisions, made apparent through changes in response times.

General Hypothesis 6 (GH6): We also hypothesize that affective responses to known/judged correct and incorrect answers during training will predict changes in prospective confidence.

General Research Question 1: We will investigate if repeated hindsight experience reduces the sensitivity of subsequent retrospective confidence to decision accuracy.

General Research Question 2: Because individuals with higher metacognitive awareness tend to be less overconfident (Stankov & Lee, 2008), we investigate metacognitive awareness as a potential moderator of the effects of hindsight bias on confidence and the

effects of emotional responses in learning confidence.

The Present Dissertation

When people look back, they often overestimate the extent to which they could have known an outcome or answer that they now know. Put differently, people are overconfident that they would have been correct when they know outcome information relative to their confidence when outcome information is unknown (Fischhoff, 1975; Winman et al., 1998). Although hindsight bias is often evaluated as a retrospective judgment of how a person would have performed (similar to retrospective confidence), the link between hindsight bias and confidence for subsequent, related tasks has been little evaluated, with no definitive evidence published about their causal relationship to the best of my knowledge. This gap is problematic because hindsight bias may influence *how* people make similar decisions in the future through changes in prospective confidence and efficacy beliefs. For example, prospective confidence may motivate persistence and more deliberative processing (Peters & Shoots-Reinhard, 2022). In addition, changes to prospective confidence may further influence retrospective confidence in subsequent tasks.

The goals of the present dissertation are to: 1) provide evidence of the causal relationship between hindsight bias and increases in both prospective and retrospective confidence, 2) study how the amount of hindsight experience can affect changes in confidence, 3) begin to understand the consequences of elevated confidence due to hindsight bias on decision processes, and 4) investigate possible emotional processes underlying how people learn confidence from experiences.

The present research is novel for a couple of reasons. First, prior research has not empirically shown a causal link between hindsight bias and effects on subsequent decision making and confidence. We seek to test the longstanding speculation that hindsight causes overconfidence for subsequent decisions (e.g., Dawson et al., 1993). Secondly, the role of prospective confidence and self-efficacy in decision making has largely been studied in perceptual decision tasks (e.g., Van Marcke et al., 2024; Boldt et al., 2019) or the cumulative effects of many decisions (e.g., Peters et al., 2019). We seek to investigate the role of prospective confidence on decision making and retrospective confidence at the level of an individual decision and to extend beyond perceptual tasks. Lastly, we seek to investigate the role of affect in learning confidence. Prior research has limited the study of affect informing confidence to incidental sources (e.g., Fieker et al., 2016; Koellinger & Treffers, 2015). We investigate affect as a critical part in how people learn how good/bad they are at a task.

To foreshadow our results, we find mixed support for hindsight bias increasing prospective and retrospective confidence relative to foresight. Despite confirmation of experienced hindsight bias (hindsight participants believing they would have answered more questions correctly than foresight participants actually answered correctly), participants' confidence *did not increase* following hindsight bias, regardless of the amount of training (Study 1). This pattern of results challenges the speculation that hindsight bias causes overconfidence. We further find that prospective confidence informs retrospective confidence judgments (Study 1), and the relation of prospective and retrospective confidence holds even while controlling for other sources of retrospective confidence (e.g., response times). Cognitive modeling suggests that hindsight bias and prospective confidence may affect the rate of evidence accumulation and cautiousness in decision making (Study 2). Thus, prospective confidence appears to play a critical role in

subsequent decision processes and retrospective confidence judgments. Finally, we found that affective responses in training are associated with prospective confidence for subsequent decisions and appear to be important predictors of confidence sensitivity and calibration (Study 2). Specifically, more positive affective responses to positive feedback appeared to increase prospective confidence. In addition, more negative affective responses to negative feedback appeared to lower subsequent retrospective confidence judgments when incorrect.

II: PILOT STUDIES

In preparation for this dissertation, we conducted three pilot studies. Each of the three pilot studies attempted to manipulate prospective confidence through three types of training experiences: foresight, foresight with feedback, and hindsight. In the foresight condition, participants answered a set of questions and were never given any outcome information. In the foresight-with-feedback condition, participants answered the same set of questions but were then told the correct answer. In effect, this training experience replicates when people make a prediction and then are told the correct answer. Hindsight may have occurred to some degree (particularly when the answer was confirmed; Bradfield & Wells, 2005), but participants also received information about their performance and ability at the task (Hoch & Loewenstein, 1989). This condition acts as a control for both the foresight condition—in that participants actually answer the questions—and the hindsight condition—in that participants learn what the correct answer is. In the hindsight condition, participants were immediately told the correct answers to the same set of questions and asked what they would have picked. This condition was most similar to that of the physicians performing right-heart catheterization (Dawson et al., 1993) in that participants never made a prediction but received outcome information. The goals of the pilot studies were to investigate if we could manipulate prospective confidence for a novel math decision task by manipulating how participants gain experience and to develop stimuli for Study 1 and Study 2.

Pilot Study 1

The first was an online pilot using Amazon's Mechanical Turk (N = 192). The goal of this first pilot was to test whether repeated experiences with a novel math task could affect prospective

confidence. We chose to specifically use math because of the availability of existing measures of numeric ability (e.g., Silverstein et al., 2023). We expected that our hindsight condition would cause participants to report that they would have been more accurate than participants actually were accurate in the foresight condition. Further, we hypothesized that repeated hindsight experiences in the hindsight condition would increase prospective confidence relative to the foresight condition (GH1).

Method

Procedure

Participants were asked to decide which of two math expressions was larger (see Figure 2.1 for an example choice). Participants were randomly assigned to gain experience in this task in one of three conditions: foresight (N = 63), foresight with feedback (N = 66), and hindsight (N = 63). All participants saw the same 15 expression pairs in a training session, with each pair shown for 5 seconds. In the *foresight condition*, participants were shown the expression pairs and were asked to indicate the larger value to the best of their ability (the correct answers were never indicated).

A **B**
OR
12% of \$81 **31% of \$37**

Figure 2.1 Example of a choice between two expressions. Participants had 5 seconds to determine the larger expression.

In the *foresight-with-feedback condition*, participants were shown the expression pairs and were asked to indicate the larger value to the best of their ability, and, after each trial, they were shown the correct response by circling the larger expression in red (see Figure 2.2). In the *hindsight condition*, the larger expression was circled in red immediately, and participants were asked to indicate which option they would have picked as the larger expression had the answer not been indicated. After the training trials, participants were asked to indicate their retrospective confidence for the training and their prospective confidence for similar trials.



Figure 2.2. Example of a choice between two expressions with the larger expression indicated. This screen was shown in foresight-with-feedback trials after each participant's response and was shown in hindsight trials immediately and before they indicated which option they would have chosen.

Measures

Choices. As a manipulation check, we summed the number of correct responses in the 15 training trials for the foresight and foresight-with-feedback conditions, as well as the number of responses participants indicated they would have been correct in the hindsight condition to assess if we produced a hindsight bias. Hindsight bias was confirmed if participants indicated the

correct answer more often in the hindsight condition than in the foresight condition (Hoch & Loewenstein, 1989).

Retrospective Training Confidence. As another manipulation check and following the training trials that featured the manipulation, participants were asked to indicate how many of the trials they believed they answered (“How many of the 15 training trials do you think you answered correctly;” in the foresight and foresight-with-feedback conditions) or would have answered correctly (“How many of the 15 training trials do you think you would have answered correctly;” in the hindsight condition) as a retrospective judgment of performance (Schraw, 2009).

Participants indicated their confidence using a slider from 0 to 15. If participants in the hindsight condition reported greater confidence in how they would have performed compared to participants' retrospective confidence in how they actually performed in the foresight condition, it further suggested a successful hindsight manipulation (Fischhoff, 1975; Hoch & Loewenstein, 1989).

Prospective Confidence. Participants were asked to give a prospective judgment of performance, “If you were to now complete 45 more trials without being given the answers, how many trials do you think you would answer correctly?” Participants indicated their confidence using a slider from 0 to 45.

Results and Discussion

Hindsight Manipulation Checks—Choices and Training Retrospective Confidence.

Participants in the hindsight condition ($M = 12.19$, $SD = 2.81$) indicated that they would have picked the correct answer more often than participants in the foresight ($M = 8.98$, $SD = 1.97$, $p_{Tukey} < .001$, $d = 1.35$) and foresight-with-feedback conditions ($M = 9.88$, $SD = 2.26$, $p_{Tukey} <$

.001, $d = 0.98$) actually picked the correct answer ($F(2, 189) = 30.82, p < .001, \eta^2 = .25$).

Participants in the foresight and foresight-with-feedback conditions did not differ significantly in their accuracy ($p = .08, d = 0.38$). See Figure 2.3.

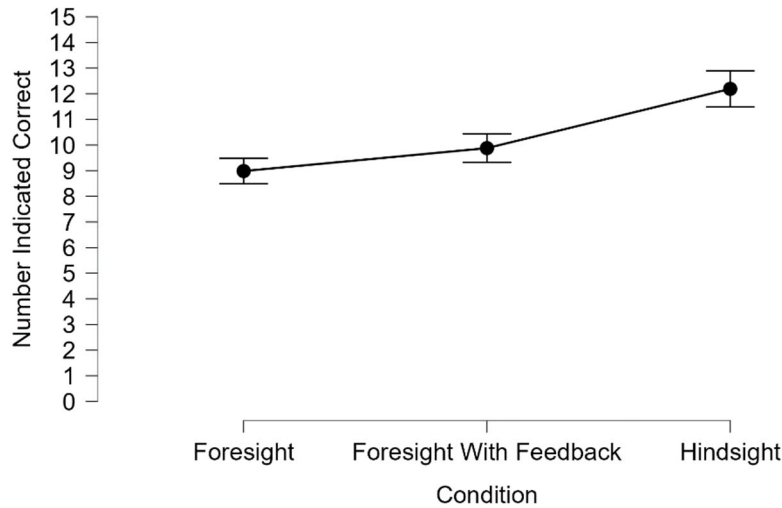


Figure 2.3. Participants' accuracy by training condition. In the hindsight condition, accuracy was determined by whether participants indicated they would have picked the correct expression. Error bars represent 95% confidence intervals.

As expected, participants in the hindsight condition ($M = 9.79, SD = 2.82$) indicated greater retrospective confidence in how they would have performed compared to participants' retrospective confidence in how they performed in the foresight condition ($M = 7.49, SD = 2.69, p_{Tukey} < .001, d = 0.82$). Participants in the foresight-with-feedback condition ($M = 9.33, SD = 2.82$) also indicated greater confidence in their ability retrospectively than participants in the foresight condition ($p_{Tukey} < .001, d = 0.66, F(2, 189) = 11.98, p < .001, \eta^2 = .11$). Participants in the hindsight and foresight-with-feedback conditions did not significantly differ in their training retrospective confidence ($p = .62, d = 0.17$). See Figure 2.4. The results for both participants' choices and training retrospective confidence suggest that our novel paradigm did produce repeated hindsight bias among participants in the hindsight condition, consistent with previous findings (Fischhoff, 2007; Hoch & Loewenstein, 1989). Relative to participants in the foresight

condition, participants in the hindsight condition chose the correct answer more often and reported greater retrospective confidence.

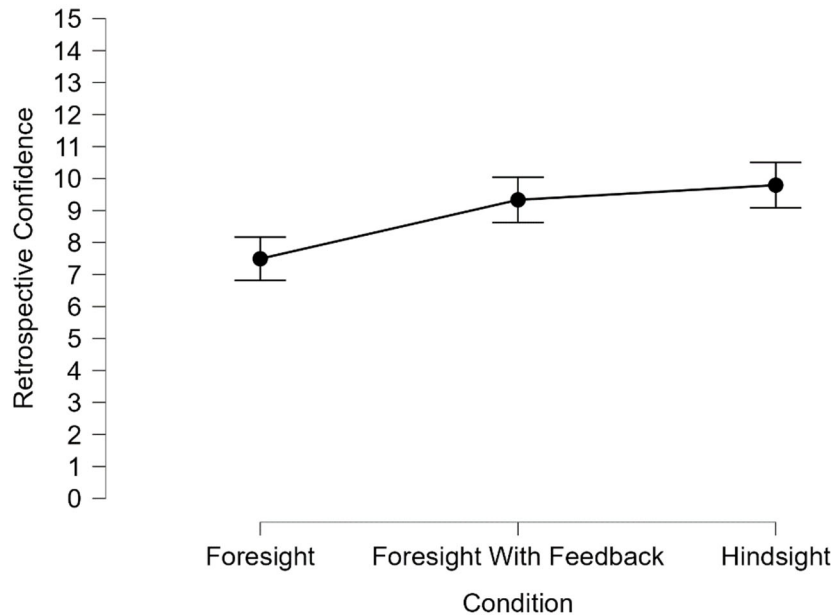


Figure 2.4. Training retrospective confidence by training condition. Error bars represent 95% confidence intervals.

Hindsight and Prospective Confidence. As hypothesized (GH1), participants in the hindsight condition ($M = 27.54$, $SD = 7.86$) indicated greater confidence in their ability prospectively than participants in the foresight condition ($M = 23.46$, $SD = 8.97$, $p_{Tukey} = .02$, $d = 0.49$). Participants in the foresight-with-feedback condition ($M = 29.17$, $SD = 8.39$) also indicated greater prospective confidence than participants in the foresight condition ($p_{Tukey} < .001$, $d = 0.68$, $F(2, 189) = 7.81$, $p < .001$, $\eta^2 = .08$). Participants in the hindsight and foresight-with-feedback conditions did not significantly differ in their prospective confidence ($p = .52$, $d = 0.19$). See Figure 2.5. This finding is novel and supports the speculation of Dawson and colleagues (1993).

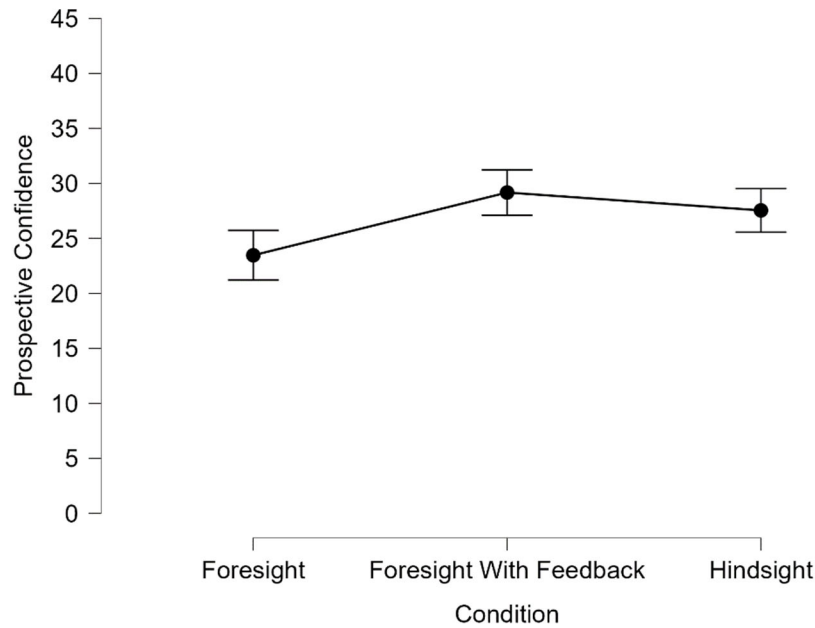


Figure 2.5. Prospective confidence by training condition. Error bars represent 95% confidence intervals.

Pilot Study 1 Summary

The hindsight manipulation sought to afford participants repeated hindsight experiences. As we hoped, participants did appear to experience hindsight bias in the hindsight condition. Relative to participants in the foresight condition, participants in the hindsight condition chose the correct answer more often and reported greater training retrospective confidence consistent with past hindsight experiments (e.g., Hoch & Loewenstein, 1989). Additionally, repeated hindsight experiences elevated participants' prospective confidence relative to the foresight condition (GH1). The increase in prospective confidence is consistent with retrospective confidence's function in learning prospective confidence as detailed in the Confidence Cycle (see Chapter 1 for more information). These findings go beyond the findings of Bradfield and Wells (2005)—which suggested that hindsight bias following a prediction's confirmation increases efficacy beliefs—because participants in the hindsight condition made no predictions to be confirmed or

rejected. As such, the outcome information in the hindsight condition was spurious as to whether participants would have been accurate and is not inherently informative about participants' abilities. It also directly tests the hypothesized causal relationship between repeated hindsight experiences and prospective confidence speculated by Dawson and colleagues (1993). However, this pilot did not assess whether repeated hindsight experiences would 1) bias subsequent retrospective confidence judgments (GH2) or 2) influence decision processes in similar, subsequent choices (GH5). Nor did it vary the amount of experience to investigate the effects of more or less hindsight experiences on prospective confidence and retrospective confidence (GH3 and GH4). In Pilot 2, we begin to test whether our methodology is capable of assessing our hypothesis regarding subsequent decisions (GH2 and GH5).

Pilot Study 2

The primary goal of this pilot was to assess whether the same stimuli would be effective for an in-person student sample. An in-person sample will be necessary for the collection of physiological data and clean response time data for cognitive modeling as part of Study 2. A secondary goal of this study was to explore the effects of hindsight-induced prospective confidence on decision-making processes and trial-by-trial retrospective confidence for subsequent decisions. We expected that our hindsight condition would cause a hindsight bias relative to the foresight condition. We hypothesized that repeated hindsight experiences in the hindsight condition would increase prospective confidence relative to the foresight condition (GH1). We expected that hindsight-induced increases in prospective confidence would cause changes in participants' response times (GH5) for similar trials and increase trial-by-trial retrospective confidence (GH2). We were interested in how the sensitivity of trial-by-trial retrospective confidence would be impacted by the hindsight condition (GRQ1). We also

explored how prospective confidence, response times, objective accuracy, and numeracy predicted trial-by-trial retrospective confidence. We expected retrospective confidence to be greater for correct responses than incorrect responses (Ariely et al., 2000). We expected that numeracy might interact with accuracy, given previous findings that confidence relies on the same evidence accumulation and ability as decisions (Dunning et al., 2003; Pleskac & Busemeyer, 2010). Longer response times were expected to be associated with less retrospective confidence (Pleskac & Busemeyer, 2010).

Method

Procedure

The second pilot was in person, using the PsychLing Subject Pool at the UO (N = 100 students). First, all participants completed baseline measures of numeracy and subjective numeracy. Then, participants were asked to decide which of two math expressions was larger, similar to Pilot Study 1. Participants were randomly assigned to gain experience in this task in one of three ways: foresight (N = 34), foresight with feedback (N = 33), and hindsight (N = 33). As in Pilot Study 1, all participants saw the same 15 expression pairs in this training session, with each pair shown for 5 seconds. In the *foresight* condition, participants were shown the expression pairs and were asked to indicate the larger-value expression to the best of their ability (the correct answers were never indicated). In the *foresight-with-feedback* condition, participants were shown the expression pairs and were asked to indicate the larger-value expression to the best of their ability, and, after each trial, they were shown the correct response by circling the larger expression in red. In the *hindsight* condition, the larger expression was circled in red immediately, and

participants were asked to indicate which option they would have picked as the larger expression had the answer not been indicated.

After the training trials, participants completed measures of training retrospective confidence and prospective confidence for the trials. Then, all participants completed 45 test trials without any feedback. Following each trial, participants were asked to indicate their confidence in their immediately prior answer as a measure of trial-by-trial retrospective confidence.

Measures

Objective Numeracy. Numeracy was assessed using the Adaptive Numeric Understanding Measure, a measure that requires each participant to complete four math items that vary based on the participant's performance (Silverstein et al., 2023). All participants saw the same initial item: "Imagine that you have a 5-sided die (the sides of which show 1, 2, 3, 4, and 5), and we throw it 150 times. On average, out of these 150 throws how many times would this 5-sided die show an odd number (1, 3, and 5)? _____ throws." Ultimately, the measure categorized participants into nine levels of numeracy.

Subjective Numeracy. This 8-item measure from Fagerlin and colleagues (2007) has 2 subscales: numeric confidence (e.g., 'How good are you at working with fractions?') and preference for numeric information (e.g., 'How often do you find numerical information to be useful?'). Participants responded on 6-point Likert-type scales. Subjective numeracy was calculated as an average of all items. Numeric self-efficacy was calculated as the average of the first 4 items, and numeric preference as the average of the last 4 items.

Training Retrospective Confidence. After the training trials, participants were asked to indicate how many of the trials they believed that they answered/would have answered correctly, as a measure of training retrospective confidence. Participants in the foresight and foresight-with-feedback conditions were asked, “How many of the training trials do you think you got correct?” Participants in the hindsight condition were asked, “How many of the training trials do you think you would have gotten correct?” Participants responded using a slider from 0-15. This question was used as a manipulation check to ensure that the hindsight condition induced a hindsight bias. If participants in the hindsight condition reported greater confidence in how they would have performed compared to participants' retrospective confidence in how they actually performed in the foresight condition, it suggested a successful hindsight manipulation (Fischhoff, 1975; Hoch & Loewenstein, 1989).

Prospective Confidence. Similar to Boldt and colleagues (2019), participants were asked to indicate their prospective confidence for the 45 test trials. They were asked, “Before you get started, how many of the following 45 trials do you think that you will answer correctly?” Participants responded using a slider from 0-45. For regression analyses, the value provided was transformed into a proportion by dividing the value by 45. This transformation was done so that prospective confidence and trial-by-trial retrospective confidence would be on the same scale.

Trials-By-Trial Retrospective Confidence. Following each trial, participants were asked to indicate their confidence in their immediately prior answer on a 0 (Certain Incorrect) to 100 (Certain Correct) scale as a measure of trial-by-trial retrospective confidence. Participants were asked, “How confident are you in your previous answer?” They were instructed on how to use the slider: “Please use the slider below to indicate your belief of the probability that you are correct. For example, if you know that you didn’t respond in time and are certain you are

incorrect you would indicate 0 on the slider. If you completely guessed and have no clue, you would indicate 50 on the slider.” The value provided was treated as a proportion of 100.

Results and Discussion

Manipulation Check-Training Retrospective Confidence. Unlike in Pilot Study 1 and counter to what was expected, participants’ retrospective confidence for the 15 training trials did not significantly differ by training condition, $F(2, 97) = 1.97, p = .15, \eta^2 = .04$. See Table 2.1. One reason for this effect may be that the problems were too difficult, causing a floor effect. Another reason could be that the sample was too small. Supporting this second reason, the effect was in the same direction as in Pilot Study 1.

Descriptives – Training Retrospective Confidence

Condition	N	Mean (out of 15)	SD	SE
Foresight with Feedback	33	7.242	2.634	0.459
Foresight	34	6.147	1.987	0.341
Hindsight	33	7.242	3.123	0.544

Table 2.1. Descriptive statistics of training retrospective confidence by training condition

Trials Prospective Confidence. Also unlike Pilot Study 1, participants’ prospective confidence for the 45 test trials did not significantly differ by training condition, $F(2, 97) = 2.55, p = .08, \eta^2 = .05$. See Table 2.2.

Descriptives – Post-Training Trials Prospective Confidence

Condition	N	Mean (out of 45)	SD	SE
Foresight with Feedback	33	21.424	8.109	1.412
Foresight	34	17.176	8.159	1.399
Hindsight	33	20.818	8.777	1.528

Table 2.2. Descriptive statistics of prospective confidence by training condition

Prospective Confidence and Response Times. Given that there were no significant condition differences in prospective confidence, we turned to a correlation exploration of the relation between prospective confidence and response times. A hierarchical linear regression was fitted predicting response times in the trials by participants' prospective confidence with random intercepts for each participant and each expression pair. Consistent with findings from Boldt and colleagues (2019) and our hypotheses, participants who reported greater prospective confidence tended to respond more slowly ($b = 0.04, p = .02$). See Figure 2.6. This association suggests that prospective confidence may have motivated more deliberation when answering the 45 test trials. Hindsight-induced increases in prospective confidence may similarly increase deliberation (GH5).

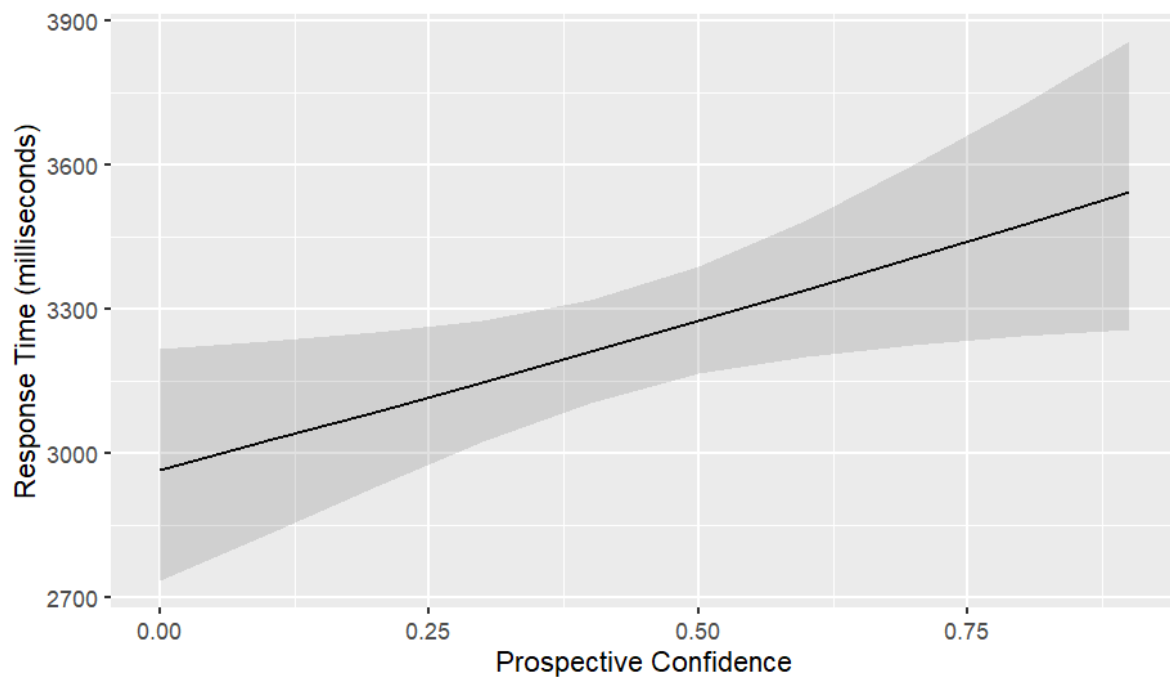


Figure 2.6. Predicted marginal effect of prospective confidence on trial-by-trial response times

Prospective Confidence and Trial-by-Trial Retrospective Confidence. Hierarchical linear regression was fitted to predict trial-by-trial retrospective confidence with participants'

prospective confidence and random intercepts for each participant and each expression pair. Participants who reported greater prospective confidence tended to be more confident in their answers ($b = 2.67, p = .008$). See Figure 2.7. This association is consistent with the Confidence Cycle and the function of prospective confidence in informing retrospective confidence (Ehrlinger & Dunning, 2003). Although not a direct test of GH2—that hindsight-induced increases in prospective confidence would increase trial-by-trial retrospective confidence—, this finding suggests that repeated hindsight experience could inflate confidence for subsequent, similar tasks.

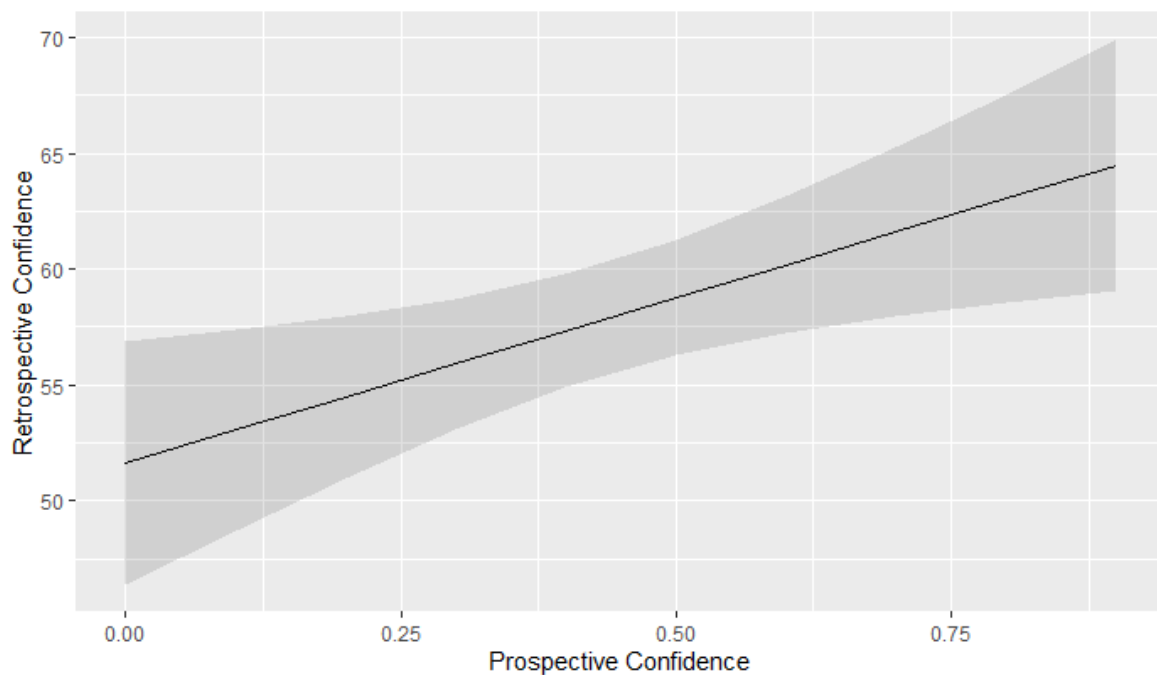


Figure 2.7. Predicted marginal effect of prospective confidence on trial-by-trial retrospective confidence

It is plausible that participants with greater prospective confidence had greater retrospective confidence because they were actually correct more often. To investigate whether accuracy can explain the relation between prospective and retrospective confidence, another hierarchical linear regression was fitted predicting trial-by-trial retrospective confidence by participants'

prospective confidence, controlling for trial-by-trial accuracy with random intercepts for each participant and each expression pair. Participants who reported greater prospective confidence tended to be more confident in their answers ($b = 2.40, p = .02$), and participants also tended to be more confident in their answers when they were correct v. incorrect ($b = 9.20, p < .001$). See Figure 2.8.

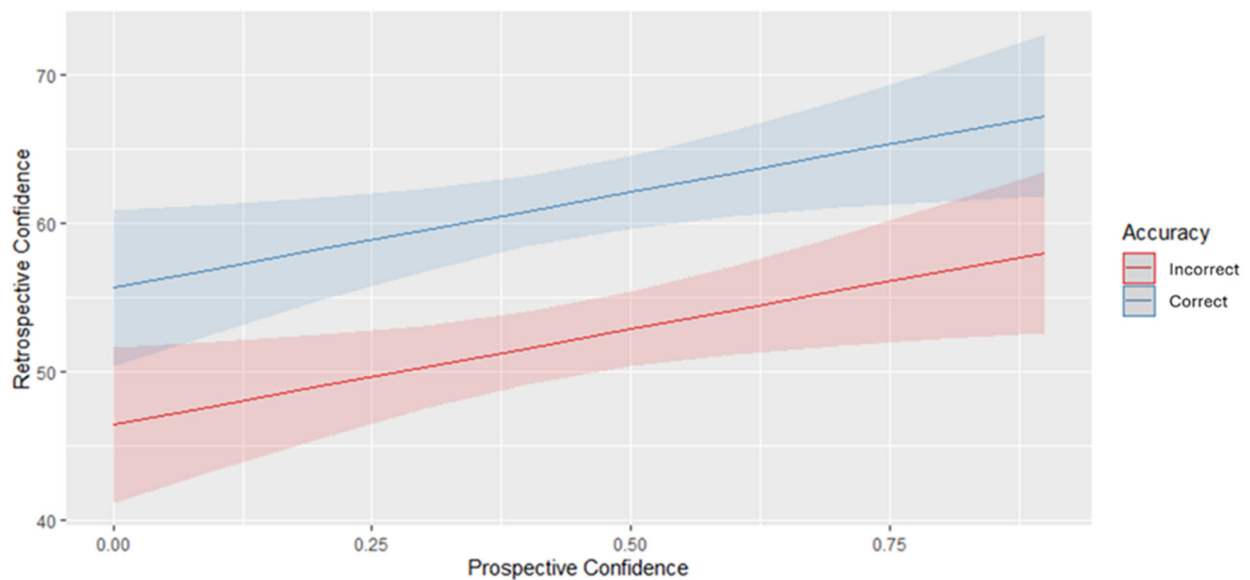


Figure 2.8. Predicted marginal effects of prospective confidence and trial-by-trial accuracy on trial-by-trial retrospective confidence

To explore what else predicts trial-by-trial retrospective confidence, a final hierarchical linear regression was fitted predicting trial-by-trial retrospective confidence with participants' prospective confidence (standardized), response times (log-transformed and standardized), trial-by-trial accuracy, numeracy (standardized), interactions of these variables, and random intercepts for each participant and each expression pair. Given the exploratory nature of this analysis, we then pared down the model to remove non-significant, higher-order interactions. Prospective confidence appeared to interact with response times ($b = -1.04, p = .001$), such that prospective confidence appears to inform retrospective confidence (consistent with prospective confidence

informing retrospective confidence and the proposed Confidence Cycle); however, the weight of prospective confidence in retrospective confidence was reduced when the decision took longer. This interaction appears to be consistent with Koriat (2007), who suggested that retrospective confidence was informed by multiple heuristic cues. See Figure 2.9. Additionally, accuracy appeared to interact with numeracy ($b = -2.56, p < .001$), such that highly numerate participants' retrospective confidence was more sensitive to their accuracy, whereas the less numerate had about equally low confidence after a trial whether they answered it correctly or not. See Figure 2.10.

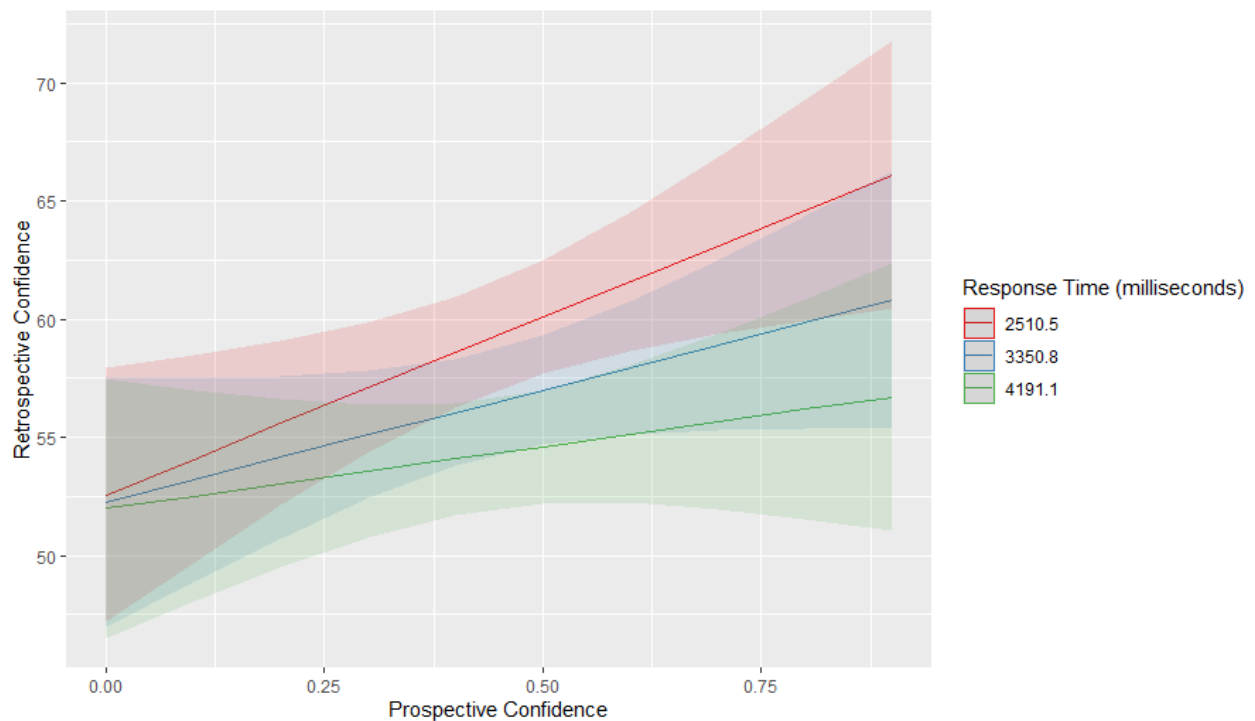


Figure 2.9. Predicted marginal effects of prospective confidence and trial-by-trial response times on trial-by-trial retrospective confidence

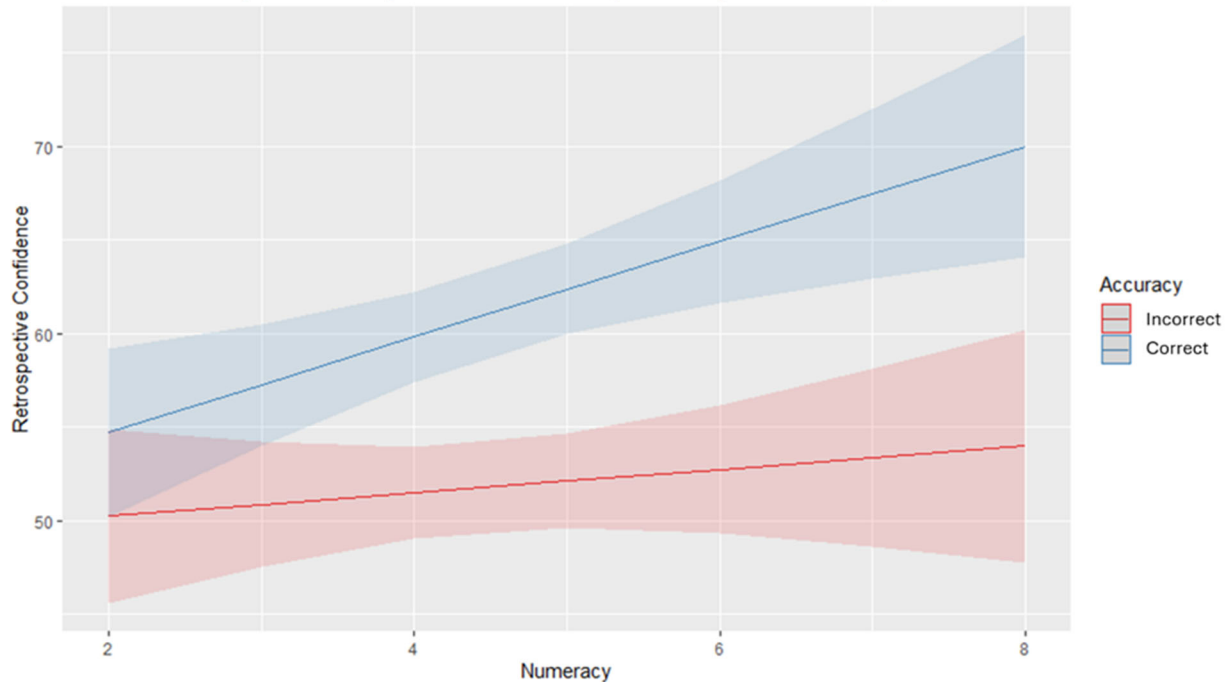


Figure 2.10. Predicted marginal effects of numeracy and trial-by-trial accuracy on trial-by-trial retrospective confidence

Pilot Study 2 Summary

The hindsight manipulation was not significant in this pilot. One reason might be that the expression pairs were very difficult for an undergraduate psychology and linguistics student population. Another reason may be that Pilot 2 was underpowered to find significant differences. Consistent with the latter explanation, the effect of hindsight was in the predicted direction. Nonetheless, we reasoned that we could increase the effect of the manipulation in the next pilot study by making the items easier and increasing the number of training trials. Consistent with Boldt and colleagues (2019), greater prospective confidence was associated with slower decision making. This association could suggest that more confident participants were motivated to deliberate for longer. However, we cannot conclude that prospective confidence caused slower decision making due to the correlational nature of the variables. Nor can we conclude that the nature of the relation between prospective confidence and response times will be the same

following repeated hindsight experiences. This finding is consistent with the Confidence Cycle and the function of prospective confidence in motivating deliberation. It is inconsistent with findings that suggest prospective confidence might demotivate deliberation (Alter et al., 2007; Bandura & Jourden, 1991; Dutilh et al., 2012; Vancouver et al., 2002). If this pattern holds for hindsight-induced increases in prospective confidence, it suggests that hindsight causes people to deliberate more. For people high in ability, increased confidence may benefit them (Peters, Tompkins, et al., 2019). However, for those lower in ability, increasing confidence may be detrimental.

Numeracy related to recognizing correct responses in retrospective confidence judgments. This finding is consistent with the idea that confidence and choice are based on the same evidence accumulation processes and skills (Kruger & Dunning, 1999; Pleskac & Busemeyer, 2010). As expected, prospective confidence appeared to inform retrospective confidence judgments, particularly for quick decisions. When decisions took longer, time may have acted as a heuristic cue for difficulty, leading people to reduce their confidence (Koriat, 2007; Pleskac & Busemeyer, 2010). This additional cue appeared to have “watered down” the informational value of prospective confidence. Due to the failure of the training manipulation to change prospective confidence, we cannot speak to the causal relationship between prospective and retrospective confidence. However, the correlational relation between prospective and retrospective confidence is consistent with the Confidence Cycle and the functions of prospective confidence.

Pilot Study 3

We hypothesized that the effects of hindsight on prospective confidence and retrospective confidence in subsequent decisions would increase with more experience (GH3 and GH4).

Additionally, we sought to make the stimuli easier than in Pilots 1 and 2 to improve the power of the manipulation. Consistently, intending to increase power, we collected a larger sample in this final pilot. Thus, the goal of the third pilot study was to identify stimuli to use for the proposed studies and determine training phase lengths. The trials used in this pilot were all easier than those used in Pilots 1 and 2, despite varying in difficulty as an experimental condition. Two questions remain: 1) Should we use easy or hard trials in the later dissertation studies, and 2) does a longer training phase length increase confidence over a shorter length (GH3 and GH4)?

Method

We ran an online study using Amazon's Mechanical Turk (N = 309). As in Pilot Studies 1 and 2, participants were asked to decide which of two math expressions was larger. First, participants completed a measure of metacognitive awareness. Before the training trials but after the introduction to the task, participants were asked how many of the training trials they thought they'd answer correctly as a measure of baseline prospective confidence. Then, they were randomly assigned to gain experience in this task in one of three ways: foresight (N = 102), foresight with feedback (N = 104), and hindsight (N = 103). They were also randomly assigned to see items from one of two sets of expression pairs: difficult items or easy items. Additionally, participants were randomly assigned to experience either eight or 25 training trials (see Table 2.3 for the full breakdown of the number of participants per condition). Each pair was shown for 5 seconds. In the *foresight* condition, participants were shown the expression pairs and were asked to indicate the larger-value expression to the best of their ability (the correct answers were never indicated). In the *foresight-with-feedback* condition, participants were shown the expression pairs and were asked to indicate the larger-value expression to the best of their ability, and, after each trial, they were shown the correct response by circling the larger expression in red. In the

hindsight condition, the larger expression was circled in red immediately, and participants were asked to indicate which option they would have picked as the larger expression had the answer not been indicated.

Difficulty	Condition	Length		Total
		Long (25)	Short (8)	
Easy (~90%)	Foresight	38	21	59
	Foresight w/ Feedback	26	27	53
	Hindsight	16	28	44
	Total	80	76	156
Hard (~70%)	Foresight	21	22	43
	Foresight w/ Feedback	30	21	51
	Hindsight	24	35	59
	Total	75	78	153
Total	Foresight	59	43	102
	Foresight w/ Feedback	56	48	104
	Hindsight	40	63	103
	Total	155	154	309

Table 2.3. Sample sizes (*n*) by condition

After the training trials, participants indicated their training retrospective confidence and post-training prospective confidence for 45 more hypothetical trials.

Measures

Metacognitive Awareness. Participants answer a shortened version of the metacognitive awareness scale (Schraw & Dennison, 1994). The measure included 18 questions across the monitoring (e.g., “I consider several alternatives to a problem before I answer”; $\alpha = .79$), evaluation (e.g., “I know how well I did once I finish a test”; $\alpha = .77$), and declarative knowledge (e.g., “I understand my intellectual strengths and weaknesses”; $\alpha = .59$) subscales. Participants

responded on a 5-point Likert scale from Strongly Disagree to Strongly Agree. Items from the three subscales were averaged together.

Baseline Prospective Confidence. Before the training trials but after the introduction to the task, participants were asked how many of the training trials they thought they'd answer correctly as a measure of baseline prospective confidence, "Before you begin the training, how many of the [8/25] training trials do you think you will answer correctly?" Participants indicated their confidence using a slider from 0 to 8 or 25, depending on the length condition. The value provided was transformed into a proportion by dividing the value by the length of the training.

Choices. As a manipulation check, we summed the number of correct responses for the foresight and foresight-with-feedback conditions, as well as the number of responses participants indicated they would have given that would have been correct in the hindsight condition and divided the numbers by the length of the training to assess if we produced a hindsight bias. Hindsight bias was confirmed if participants indicated they would have picked the correct answer more often in the hindsight condition than participants selected the correct answer in the foresight condition (Hoch & Loewenstein, 1989).

Training Retrospective Confidence. As another manipulation check and following the training trials that featured the manipulations, participants were asked to indicate how many of the trials they believed they answered correctly ("How many of the [8/25] training trials do you think you answered correctly;" in the foresight and foresight-with-feedback conditions) or would have answered correctly ("How many of the [8/25] training trials do you think you would have answered correctly;" in the hindsight condition) as a retrospective judgment of performance (Schraw, 2009). Participants indicated their confidence using a slider from 0 to 8 or 25,

depending on the length condition. If participants in the hindsight condition report greater confidence in how they would have performed than participants' confidence in how they actually performed in the foresight condition, it suggests a successful hindsight manipulation (Fischhoff, 1975; Hoch & Loewenstein, 1989). For comparability, the value provided was transformed into a proportion by dividing the value by the number of training trials.

Post-Training Prospective Confidence. Participants were asked to give a prospective judgment of performance for 45 hypothetical test trials, “If you were to now complete 45 more trials without being given the answers, how many trials do you think you would answer correctly?” Participants indicated their confidence using a slider from 0 to 45. The value provided was transformed into a proportion by dividing the value by 45.

Training Retrospective Confidence Bias. We calculated participants' confidence bias for their training retrospective confidence by taking the difference in their accuracy from their training retrospective confidence as a measure of over/under-confidence (Yates, 1982). This operationalization results in a measure between -1 and 1, where positive values indicate overconfidence and negative values indicate underconfidence.

Results and Discussion

Hindsight Manipulation Check. Length of the training didn't have a simple effect on the proportion of correct answers indicated, nor did it interact with the training condition and the difficulty. This null effect was expected. Even with the amount of hindsight bias proportionately the same across the lengths, more experiences will result in more spuriously positive experiences in hindsight. For the easy trials, participants in the hindsight condition ($M = 0.93$, $SD = 0.15$) did not indicate that they would have picked the correct answer more often than participants in the

foresight condition ($M = 0.88$, $SD = 0.18$, $d = 0.31$, $p_{Tukey} = .63$) and foresight-with-feedback conditions ($M = .93$, $SD = 0.15$, $d = 0.04$, $p_{Tukey} > .99$) picked the correct answer. Thus, the hindsight manipulation check in the easy trials was unsuccessful.

For the hard trials, participants in the hindsight condition ($M = 0.86$, $SD = 0.17$) indicated that they would have picked the correct answer more than participants in the foresight ($M = 0.69$, $SD = 0.16$, $d = 1.01$, $p_{Tukey} < .001$) and foresight-with-feedback conditions ($M = .74$, $SD = 0.19$, $d = 0.73$, $p_{Tukey} = .002$) actually picked the correct answer. The foresight and foresight-with-feedback conditions did not significantly differ from each other ($d = 0.28$, $p_{Tukey} = .75$; see Figure 2.11). Thus, given a presumed ceiling effect for the easy items, the harder items appeared better suited for the hindsight manipulation in our later proposed studies. These results are consistent with findings that more difficult problems produce stronger hindsight bias (Winman et al., 1998; Hoch & Loewenstein, 1989). The remaining analyses will focus only on the participants in the hard conditions.

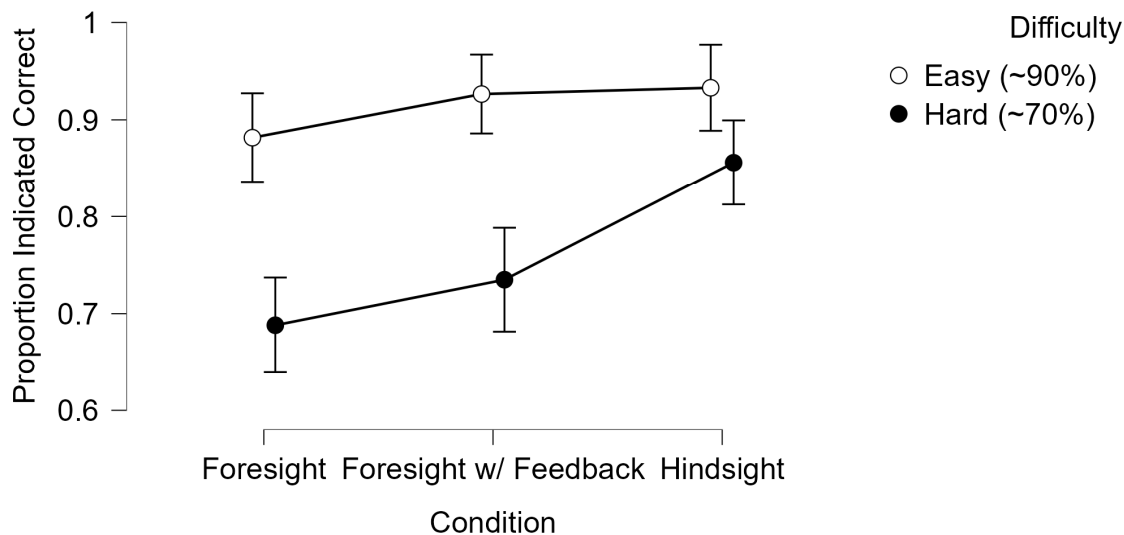


Figure 2.11. Participants' accuracy by training and difficulty conditions. In the hindsight conditions, accuracy was determined by whether participants indicated they would have picked the correct expression. Error bars represent 95% confidence intervals.

Hindsight and Change in Prospective Confidence. The change in prospective confidence was calculated as the difference between post-training prospective confidence and baseline prospective confidence. A 3 (training condition) x 2 (training length) ANOVA was run to test for differences in the change in prospective confidence as a result of the training trials. There was a significant effect of training condition ($F(2, 147) = 6.92, p = .001, \eta^2 = .08$), such that participants had a greater increase in prospective confidence in the hindsight ($p_{Tukey} = .02, d = 0.55$) and foresight-with-feedback ($p_{Tukey} = .001, d = 0.75$) compared to foresight. There was no significant difference between hindsight and foresight-with-feedback. There was no main effect of training length ($F(1, 147) = 0.51, p = .48, \eta^2 = .003$). As expected, there was a significant interaction between training condition and length ($F(2, 147) = 4.47, p = .01, \eta^2 = .05$).

In the long training phase conditions, participants in the hindsight condition ($M = .07, SD = 0.18$) indicated a marginally greater change in prospective confidence than participants in the foresight condition ($M = -0.06, SD = 0.15, p_{Tukey} = .06, d = 0.85$); however, this difference did not attain significance. Participants in the foresight-with-feedback condition ($M = 0.01, SD = .14$) did not indicate a greater change in prospective confidence than participants in the foresight condition ($p_{Tukey} = .56, d = 0.47$). Participants in the hindsight and foresight-with-feedback conditions did not significantly differ in their prospective confidence change ($p = .74, d = 0.38$). See Figure 2.12.

In the short training phase conditions, participants in the foresight ($M = -0.07, SD = .14$) and hindsight conditions did not differ ($M = -0.04, SD = 0.14, p_{Tukey} = .94, d = 0.25$). Participants in the foresight-with-feedback condition ($M = 0.08, SD = 0.10$) indicated a greater change in prospective confidence than participants in the foresight condition ($p_{Tukey} = .01, d = 1.04$) and marginally greater change than participants in the hindsight conditions ($p_{Tukey} = .06, d = 0.79$).

There was no difference in the change in prospective confidence between the long and short conditions for the foresight conditions ($p_{Tukey} = .99, d = 0.11$), nor for the foresight-with-feedback conditions ($p_{Tukey} = .60, d = 0.46$). There was a marginal difference in the change in prospective confidence between the long and short conditions for the hindsight condition ($p_{Tukey} = .09, d = 0.70$).

This pattern of results suggests that, in hindsight, participants in the long condition may have experienced more spuriously positive outcomes than in the short condition, which inflated their prospective confidence consistent with GH3. Foresight-with-feedback also increased prospective confidence relative to foresight, but this effect appears to be more related to the informative value of the feedback because more experience in foresight-with-feedback and foresight did not increase confidence beyond the shorter training condition.

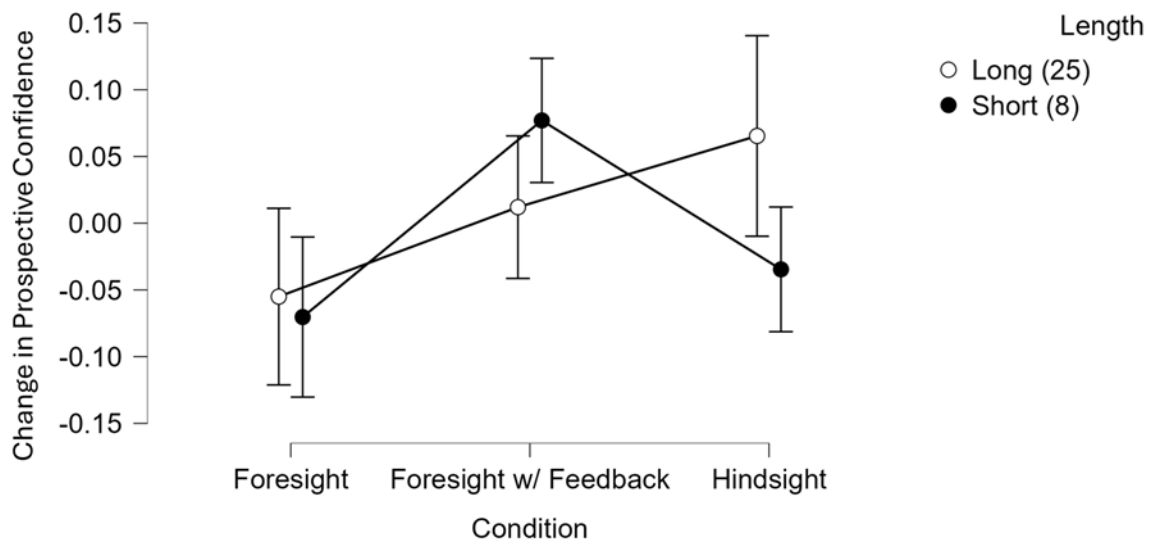


Figure 2.12. Change in prospective confidence by training and length conditions. Error bars represent 95% confidence intervals.

Metacognitive Awareness. Metacognitive awareness has previously been shown to be associated with less overconfidence (Stankov & Lee, 2008). To investigate if that relation holds

for our task, we predicted training retrospective confidence bias from the training condition interaction with each subscale of the metacognitive awareness measure in analyses of covariance. A significant interaction emerged between the evaluation subscale and the training conditions ($F(2, 147) = 3.74, p = .03, \eta^2 = .04$). There were no significant simple effects or interactions with either of the other two subscales. In the foresight condition, evaluation reduced under-confidence. Evaluation, however, was unassociated with bias in the foresight-with-feedback condition. It is possible that in the foresight-with-feedback condition, all participants were better calibrated because of the feedback. The hindsight condition cannot be interpreted meaningfully. See Figure 2.13

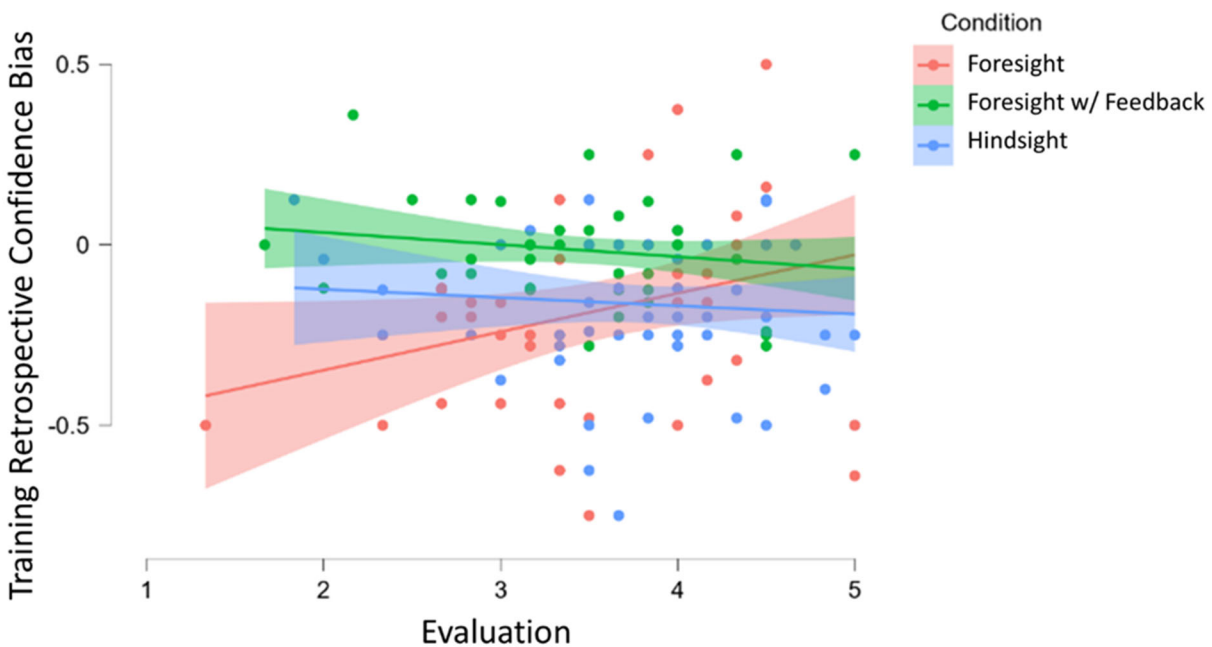


Figure 2.13. The interaction of condition and the evaluation subscale of metacognitive awareness predicting training retrospective confidence bias. Confidence bands represent 95% confidence intervals.

Pilot Study 3 Summary

The hindsight manipulation was successful in this pilot for the hard items but not for the easy items. The easy items appeared to have a ceiling effect (across all training conditions, by in large, participants picked the correct answers), suggesting they were too easy to produce a hindsight bias. Therefore, the more difficult items were used in the dissertation studies.

Additionally, the longer training phase may have produced a more robust effect of hindsight on prospective confidence. This effect of length suggests that the amount of experience may matter, particularly in hindsight. This effect, however, did not attain significance in post hoc analyses, even though the interaction was significant. This hypothesized (GH3) marginal effect is consistent with the speculation of Dawson and colleagues (1993) but has not been previously shown. With a larger sample size, we expect to see a significant difference in the effect of hindsight by training length, as well as a significant difference between the foresight and hindsight conditions following 25 training trials. Interestingly, the amount of experience mattered less for the foresight-with-feedback condition. In fact, there was less of an increase in prospective confidence directionally following more vs less experience. This pattern could be because participants experienced an initial surge in confidence that was eventually tempered by negative feedback with additional trials in the foresight-with-feedback condition. However, in the hindsight condition, the increase in trials appears to have caused an increase in confidence, possibly because the outcome is not informative about ability—outcomes are spuriously positive due to hindsight bias. In the proposed studies, we varied the amount of experience the same way we did in this pilot since the pattern of results for prospective confidence suggests there is a meaningful difference between eight and 25 training trials—there appeared to be an increase in prospective confidence in the hindsight condition but not the foresight and foresight-with-

feedback conditions. The importance of the amount of experience has important implications for the Confidence Cycle. It appears that more experiences of high retrospective confidence—even phantom confidence from spuriously positive outcomes—increase prospective confidence further. This effect is consistent with the finding that confidence was associated with experience and GH3 (Dawson et al., 1993). Finally, we began to explore metacognitive awareness (GRQ2). The evaluation subscale was associated with less underconfidence and better calibration in the foresight condition. In the present pilot, we were unable to look at the relation between retrospective confidence bias and hindsight. Further, we were not able to evaluate the role of metacognitive awareness in prospective confidence bias.

General Discussion of Pilot Studies

The three reported pilot studies have made important strides toward addressing the hypotheses, informing the design of the proposed studies, and supporting the functions of confidence as well as the Confidence Cycle. Pilot Studies 1 and 3 showed that we were successful at inducing hindsight bias, as evidenced by biased evaluations of what participants would have chosen and biased retrospective judgments of performance for the training (Fischhoff, 1975; Hoch & Loewenstein, 1989). Pilot Studies 1 and 3 further provided novel evidence that repeated hindsight experiences increase prospective confidence, supporting prior speculation (Dawson et al., 1993; GH1). These findings are consistent with retrospective confidence's function in informing prospective confidence and the Confidence Cycle. Pilot 3 confirmed prior findings that hard items produce stronger hindsight bias (Winman et al., 1998) and supported GH3 by showing that more vs less hindsight experiences increased prospective confidence. Pilot Study 3 was crucial to determining the amount of hindsight experience necessary to increase prospective confidence.

Pilot Study 2 provided correlational evidence in support of multiple hypotheses. Prospective confidence was a powerful predictor of participants' trial-by-trial retrospective confidence judgments, particularly for quick decisions. This association is consistent with prospective confidence's function in informing retrospective confidence highlighted in the Confidence Cycle (Ehrlinger & Dunning, 2003; Van Marcke et al., 2024). Further, it highlights the crucial role of the specific decision-making experience (i.e., response times) in determining retrospective confidence as well (Koriat, 2007). The result fits well with the idea that retrospective confidence is formed through a pseudo-Bayesian process with prospective confidence as a prior and decision-making experiences as data to adjust from the prior. Should this relation hold following induced hindsight bias, increases in prospective confidence should increase subsequent retrospective confidence (GH2). This finding would be the first causal evidence to our knowledge that hindsight bias increases retrospective confidence for subsequent tasks.

The Confidence Cycle suggests that prospective confidence can influence retrospective confidence both directly and through modulating decision processes. Pilot Study 2 demonstrated the correlational relation between prospective confidence and response times. The findings suggested that greater prospective confidence is associated with more cautious, effortful deliberation. Pilot 2 directionally suggests that hindsight-induced increases in prospective confidence will increase deliberation and supports GH5. However, we cannot make a causal claim about or determine the direction of the effect of repeated hindsight experiences on decision processes due to the non-significant manipulation in Pilot 2.

Although the pilot studies did not, yet, address our research question of whether metacognitive awareness would moderate the effects of repeated hindsight experience on prospective confidence, subsequent retrospective confidence, and decision processes, Pilot 3 showed that

greater metacognitive awareness—specifically evaluation tendencies—was associated with less biased retrospective confidence judgments of the training trials for participants in the foresight condition (specifically less under-confidence; GRQ2). This finding was somewhat consistent with prior findings (Stankov & Lee, 2008). The hypothesis that affective responses to outcome information will predict changes in prospective confidence (GH6) has gone unaddressed by the pilot studies but will be the target of Proposed Study 2.

III: STUDY 1: MANIPULATING PROSPECTIVE CONFIDENCE THROUGH HINDSIGHT BIAS

Study 1 aimed to examine the effects of repeated hindsight bias (the tendency to overestimate how likely you would have known something after seeing the correct answer; Fischhoff, 1975) on various metacognitive judgments and decision-making processes compared to foresight (without feedback). We attempted to manipulate prospective confidence for a numeric decision task through training trials. We were interested in whether the amount of experience in hindsight affects prospective confidence and retrospective confidence, and therefore varied both the training conditions and the length of training. Based on pilot tests, we used the hard items from Pilot 3, which the average participant answered about 70% correctly, and we varied the length to consist of either eight or 25 training trials (based on Pilot Study 3; see *Pilot Study 3* in Chapter 2 for more details). We expected that more training in hindsight would increase prospective confidence and efficacy beliefs, but not accuracy in later trials. We also examined the effects on retrospective confidence judgments and decision response times. We expected that more hindsight experience would increase retrospective confidence. We hypothesized that:

H1.1) Repeated experiences of hindsight bias would increase subsequent prospective confidence (i.e., the judged likelihood of correctness before making a decision) relative to experiences in foresight (without feedback);

H1.2) Repeated experiences of hindsight bias would increase retrospective confidence (i.e., the judged likelihood of correctness after making a decision) relative to experiences in foresight (without feedback);

H1.3) More experiences of hindsight bias would amplify the effects of repeated hindsight

experiences on prospective confidence judgments;

H1.4) More experiences of hindsight bias would amplify the effects of repeated hindsight experiences on retrospective confidence judgments;

After manipulating confidence, we investigated the effects on decision processes by examining participants' response times. Based on Pilot Study 2, we expected that a larger number of hindsight experiences would slow decision making. However, given the mixed findings in the prior literature, we also investigated a possible nonlinear relation between prospective confidence and response times. We hypothesized:

H1.5) Greater prospective confidence would be associated with slower response times for subsequent similar tasks, and, therefore, experiences of hindsight bias will slow response times relative to experiences in foresight (without feedback).

Based on the lack of association between confidence and accuracy found by Dawson and colleagues (1993), we also expected participants who experienced hindsight bias repeatedly to have less sensitive confidence judgments. We also measured metacognitive awareness since greater metacognitive awareness is associated with more accurate confidence judgments (Stankov & Lee, 2008). We explored whether repeated hindsight bias would reduce the sensitivity of retrospective confidence to objective accuracy (RQ1.1) and if individual differences in metacognitive awareness would moderate any observed effects (RQ1.2). The hypotheses, method, data cleaning procedure, and analyses for Study 1 were preregistered on the Open Science Framework (<https://osf.io/2qrbn>).

Research Design

Procedure

Study 1 consisted of a training phase (with two manipulated factors), a measurement phase, and a test phase. Before the training phase, participants completed an assessment of their metacognitive awareness (Schraw & Dennison, 1994). In the training phase, we employed a similar method and materials as in Pilot Study 3. Participants were introduced to the task through three example problems to familiarize them with responding within the time limit. Participants were then randomly assigned to one of three conditions: *hindsight* (in which participants were shown questions with the correct answers indicated and had to indicate which they would have picked), *foresight* (in which participants answered questions with no indication of the correct answer), or *foresight with feedback* (they answered questions and then were told the correct answer). The foresight-with-feedback condition acts as an experimental control for both the foresight and hindsight conditions. In the foresight-with-feedback condition, participants answer without being told the correct answer, as in the foresight condition, but they later learn the correct answer, as in the hindsight condition.

As in the pilot studies, participants in the foresight conditions—with or without feedback—had 5 seconds to choose the larger of two mathematical expressions (e.g., 25% of 30 vs. 10% of 70). Participants in the hindsight condition had 5 seconds to indicate which of two expressions they would have picked as the larger one had the larger expression not been indicated. Within each condition, participants were further assigned to one of two training lengths: short (8 trials) or long (25 trials). In the measurement phase (after the training phase and before beginning the test phase), participants reported their prospective confidence for 45 similar test trials, task efficacy

beliefs, and task difficulty beliefs. We also measured numeric self-efficacy more broadly. These measures were used to test H1.1 (that repeated hindsight experiences increase prospective confidence) and H1.3 (that more hindsight experiences further increase prospective confidence).

In the test phase, participants completed 45 trials of the same difficulty as the training trials, based on Pilot Study 2. During this phase, all participants answered the trials without feedback. After each choice, participants rated their retrospective confidence in their response. This phase was designed to assess H1.2 (that repeated hindsight experiences increase retrospective confidence), H1.4 (that more hindsight experiences further increase retrospective confidence), H1.5 (that prospective confidence slows decision making), RQ1.1 (whether hindsight experiences affect confidence sensitivity), and RQ1.2 (the moderating effect of metacognitive awareness). Following the test phase, participants completed an objective numeracy measure, and then participants in the hindsight and foresight-with-feedback conditions were asked to recall and report their experienced affect toward correct and incorrect responses in the training trials.

One week later, participants who completed the study were invited to a follow-up session to examine the persistence of the training effects. During this session, participants were reminded of the task and asked to indicate their prospective confidence, task efficacy beliefs, and task difficulty beliefs for a hypothetical 45-trial test.

Measures

Metacognitive Awareness. Participants answered a shortened version of the metacognitive awareness scale (Schraw & Dennison, 1994). The measure included 8 questions across the monitoring (e.g., “I ask myself if I have considered all options when solving a problem.”) and evaluation (e.g., “I ask myself if I have considered all options after I solve a problem.”)

subscales. Participants responded on a 5-point Likert scale from Strongly Disagree to Strongly Agree. Items from each of the two subscales were averaged to create indices of monitoring and evaluation.

Attention Checks. After an introduction to the decision task, participants were asked two 3-alternative multiple-choice questions. The first is, "Missing responses will be..." The second is, "You will have ___ seconds to respond to each trial." The answers were "considered incorrect" and "5" respectively.

Baseline Prospective Confidence. Participants were asked to indicate their prospective confidence for the 8 or 25 training trials following an introduction to the task and before the training trials. They were asked, "Now that you have some experience, we would like you to do some training for this task. We will now have you complete [8/25] training trials. Before you begin the training, how many of the [8/25] training trials do you think you will answer correctly?" Participants responded using a slider from 0 to 8 or 25, depending on which length condition they were assigned. Participants were not aware of their assigned training condition. The value provided was transformed into a proportion by dividing the value by the number of training trials.

Training Choices (Manipulation Check). As a manipulation check, we summed the number of correct responses for the foresight and foresight-with-feedback conditions, as well as the number of responses participants indicated they would have gotten correct in the hindsight condition, and divided the numbers by the length of the training to assess if we produced a hindsight bias. We would confirm that participants experienced hindsight bias if participants indicated the correct answer more often in the hindsight conditions than in the foresight conditions (without feedback;

Hoch & Loewenstein, 1989).

Post-Training Prospective Confidence. Similar to Boldt and colleagues (2019), participants were asked to indicate their prospective confidence for the 45 test trials. They were asked, “Before you get started, how many of the following 45 trials do you think that you will answer correctly?” Participants responded using a slider from 0-45. The value provided was transformed into a proportion by dividing the value by 45. This measure was repeated in the one-week follow-up for a hypothetical test of 45 items. For the follow-up, participants were asked, “Imagine we're now going to give you a test with 45 new trials of similar difficulty as your training trials from last week. How many trials do you think you would answer correctly?”

Efficacy Beliefs. To assess efficacy beliefs for this task, participants were asked, “How bad or good do you think you are at this type of task?” after having completed the training. Participants responded using a six-point semantic differential scale from “Extremely Bad” to “Extremely Good.” This measure was repeated in the one-week follow-up.

Task Difficulty Beliefs. To assess difficulty beliefs for the trials, participants were asked, “How easy or difficult do you think this version of the task is for people in general?” Participants responded using a six-point semantic differential scale from “Extremely Easy” to “Extremely Difficult.” This measure was repeated in the one-week follow-up.

Numeric Self-Efficacy. We assessed numeric self-efficacy using the ability subscale of the Subjective Numeracy Scale (Fagerlin et al., 2007; e.g., ‘How good are you at working with fractions?’). Participants responded on 6-point Likert-type scales. Responses to the four items were averaged together. This measure was repeated in the one-week follow-up.

Trials-By-Trial Retrospective Confidence. Following each trial, participants were asked to indicate their confidence in their immediately prior answer on a 0 (Certain Incorrect) to 100 (Certain Correct) scale as a measure of trial-by-trial retrospective confidence. Participants were asked, “How confident are you in your previous answer?” They were instructed how to use the slider: “Please use the slider below to indicate your belief of the probability that you are correct. For example, if you know that you didn’t respond in time and are certain you are incorrect you would indicate 0 on the slider. If you completely guessed and have no clue, you would indicate 50 on the slider.” The value provided was treated as a proportion of 100.

Overall Retrospective Confidence. Following the testing phase, participants were asked to indicate their overall retrospective confidence for the test trials. Participants were asked, “How many of the 45 trials do you think you answered correctly?” Participants indicated their responses using a slider from 0 to 45.

Affect to Outcomes. For participants in the foresight-with-feedback and hindsight conditions, we asked participants to think back to the training trials and report how they felt when they were told they were/would have been correct and incorrect separately. Affect was assessed using the valence and arousal scales of the self-assessment manikins (Bradley & Lang, 1994).

Objective Numeracy. Numeracy was assessed using the Adaptive Numeric Understanding Measure, a measure that requires each participant to complete four math items that vary based on the participant’s performance (Silverstein et al., 2023). All participants saw the same initial item: “Imagine that you have a 5-sided die (the sides of which show 1, 2, 3, 4, and 5), and we throw it 150 times. On average, out of these 150 throws how many times would this 5-sided die show an

odd number (1, 3, and 5)? _____ throws.” Ultimately, the measure sorted participants into nine levels of numeracy.

Nonsense Responding. We included two questions to check for nonsensical responding. Participants were asked, "How many days per week do you typically go swimming?" and "How many hours in a day do you typically read/watch/listen to the news?" Answers larger than 7 and 24, respectively, were considered nonsense.

Prospective Confidence Bias. Participants’ accuracy in the 45 test trials was assessed as a proportion of correct responses. The difference between reported post-training prospective confidence and accuracy on the test trials was calculated. The resulting value varied between -1 and 1, with negative values representing underconfidence and positive values representing overconfidence relative to their own accuracy.

Retrospective Confidence Brier Scores, Bias, and Sensitivity. Participants’ calibration was calculated using Brier scores on the trial-by-trial retrospective confidence judgments (Brier, 1950). Brier scores quantify the average “distance” a participant is from perfect calibration (i.e., 100% confidence when correct and 0% confidence when incorrect) We also calculated participants’ retrospective confidence sensitivity (i.e., the difference in trial-by trial retrospective confidence when correct v. incorrect; often called slope) and retrospective confidence bias (i.e., the difference between average confidence and true accuracy; often called accuracy; Yates, 1982), and prospective confidence bias (i.e., the difference between predicted and true accuracy).

Demographics. Participants were asked about their gender identity, age, race/ethnicity, education level, household income, ZIP code, and political ideology.

Data Collection

1,000 participants were recruited via Amazon's Mechanical Turk. To ensure good data quality, participants were required to have completed at least 5,000 HITs on the platform with a 99% approval rate or better, have an IP address in the United States, and be on CloudResearch's approved participants list. Participants were directed from Mechanical Turk to a Qualtrics survey link. Following the survey's completion, participants were provided a completion code for Mechanical Turk to receive compensation.

Data Cleaning

To clean the data, we first identified participants whose response times substantially differed from the average overall response time—whose overall survey response time was greater than 3 standard deviations above or less than 1.5 standard deviations below the mean. Next, we identified any participant who provided nonsense answers to open-ended responses (two questions in the demographics ["How many days per week do you typically go swimming?" and "How many hours in a day do you typically read/watch/listen to the news?"]). Any participant who responded with more than one nonsense answer was excluded. Next, we checked the attention checks. If participants answered both attention checks incorrectly, they were excluded. Additionally, if the participant substantially differed in survey time and had any nonsense answers/failed attention checks, they were excluded.

Any response greater than 3 standard deviations from the average response on a given question (aside from the trial-by-trial retrospective confidence questions) was considered missing for that question. For the trial-by-trial retrospective confidence questions, responses were considered missing if the time until the first click was the same as the page submission time, which would

indicate that participants did not attempt to answer the question. Finally, any participant missing more than 5% of the responses was excluded from any analyses. After cleaning, 842 participants were retained for data analysis, 761 of whom also completed the 1-week follow-up (90.4% retention). We were not concerned that a large number of participants were removed because this study demanded a lot of attention and focus. A logistic regression predicting retention from the training conditions, length, and interaction of condition and length indicated that participants' dropout between the initial collection and the one-week follow-up was not significantly associated with the manipulations, $\chi^2(df = 836) = 4.30, p = .51$. Our sample was majority white (75.1%) and majority male identifying (55.3%). Demographics and descriptive statistics can be found in Table 3.1.

Data Analysis

In general, participants were underconfident both prospectively and retrospectively. In the test phase, participants answered on average about 75% of the trials correctly. Participants' average prospective confidence was around 63%, and their average retrospective confidence was around 69%. Descriptive statistics can be found in Table 3.1, and bivariate correlations can be found in Table 3.2.

		Percent	Missing	Mean	Std. Deviation	Minimum	Maximum
Numeracy		-	0	4.78	1.54	1	9
Numeric Self-Efficacy		-	0	4.28	1.19	1	6
Monitoring		-	0	4.06	0.67	1	5
Evaluation		-	0	3.83	0.79	1	5
Baseline Prospective Confidence (Training)		-	0	0.64	0.19	0	1
Post-Training Prospective Confidence (Test)		-	0	0.63	0.20	0.04	1
Prospective Confidence Bias (Test)		-	0	-0.13	0.20	-0.76	0.51
Task Self-Efficacy		-	0	3.84	1.10	1	6
Task Difficulty		-	13	4.19	0.94	2.00	6
Overall Test Accuracy		-	0	0.75	0.12	0.40	1
Average Response Time (Test)		-	0	3.19	0.59	1.22	4.68
Retrospective Confidence Bias (Test)		-	0	-0.10	0.15	-0.69	0.48
Retrospective Confidence Sensitivity (Test)		-	3	0.17	0.12	-0.27	0.71
Retrospective Confidence Brier Scores (Test)		-	0	0.17	0.08	0.01	0.67
Average Trial-by-Trial Confidence (Test)		-	0	0.69	0.15	0.03	1
Overall Retrospective Confidence (Test)		-	0	0.60	0.20	0.02	1
Follow-Up Prospective Confidence		-	81	0.60	0.20	0.02	1
Follow-Up Task Self-Efficacy		-	81	3.80	1.12	1	6
Follow-Up Task Difficulty		-	81	4.22	0.92	1	6
Follow-Up Numeric Self-Efficacy		-	81	4.36	1.19	1	6
Age		-	7	45.35	12.36	20	81
Race/Ethnicity	White	75.1%	-	-	-	-	-
	Asian	9.0%					
	Black/African American	8.3%					
	Hispanic	2.9%					
	Native Hawaiian/Pacific Islander	0.1%					
	Mixed Race	4.3%					
	High School or Less	25.7%	-	-	-	-	-
Education	2-Year College	15.0%					
	4-Year College	44.8%					
	Master's Degree	12.2%					
	PhD	2.4%					
Gender	Male	55.3%	-	-	-	-	-
	Female	43.7%					

Table 3.1. Descriptive statistics of Study 1 variables and demographics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
<i>Individual Differences</i>																			
1. Numeracy																			
2. Numeric Self-Efficacy	.43***																		
3. Monitoring	-.02	.13***																	
4. Evaluation	-.12***	.11***	.75***																
<i>Training and Measurement Phases</i>																			
5. Baseline Prospective Confidence (Training)	.28***	.47***	.06	.05															
6. Post-Training Prospective Confidence (Test)	.33***	.50***	.06	.04	.74***														
7. Prospective Confidence Bias (Test)	.08*	.35***	.06	.08*	.58***	.81***													
8. Task Self-Efficacy	.24***	.54***	.09*	.11**	.58***	.69***	.52***												
9. Task Difficulty	.07	.02	.04	.03	-.16***	-.20***	-.17***	-.24***											
<i>Testing Phase</i>																			
10. Overall Test Accuracy	.41***	.24***	-.01	-.06	.25***	.31***	-.30***	.29***	-.06										
11. Average Response Time (Test)	.21***	.15***	-.06	-.05	.05	.11**	.17***	-.03	.21***	-.10**									
12. Average Trial-by-Trial Confidence (Test)	.34***	.54***	.15***	.10**	.55***	.57***	.33***	.57***	-.12***	.40***	.01								
13. Retrospective Confidence Bias (Test)	-.02	.32***	.16***	.16***	.34***	.30***	.49***	.36***	-.10**	-.31***	-.09*	.70***							
14. Retrospective Confidence Sensitivity (Test)	.46***	.27***	.02	-.08*	.17***	.18***	-.03	.09**	.08*	.34***	.19***	.26***	-.04						
15. Retrospective Confidence Brier Scores (Test)	-.54***	-.48***	-.06	.02	-.41***	-.47***	-.10**	-.39***	-.02	-.61***	-.19***	-.68***	-.17***	-.58***					
16. Overall Retrospective Confidence (Test)	.38***	.54***	.06	.04	.65***	.78***	.52***	.68***	-.17***	.43***	.05	.73***	.39***	.24***	-.58***				
<i>One-Week Follow-Up</i>																			
17. Follow-Up Prospective Confidence	.35***	.54***	.07*	.04	.65***	.77***	.55***	.66***	-.15***	.37***	.05	.68***	.39***	.20***	-.52***	.83***			
18. Follow-Up Task Self-Efficacy	.25***	.58***	.12***	.14***	.53***	.58***	.41***	.74***	-.12***	.29***	.01	.60***	.37***	.15***	-.45***	.69***	.73***		
19. Follow-Up Task Difficulty	.08*	-.03	-.02	-.04	-.16***	-.14***	-.14***	-.20***	.54***	.01	.17***	-.12***	-.15***	.11**	-.07*	-.17***	-.18***	-.23***	
20. Follow-Up Numeric Self-Efficacy	.42***	.92***	.12**	.10**	.46***	.48***	.33***	.51***	.05	.24***	.17***	.54***	.31***	.28***	-.50***	.54***	.56***	.61***	-.02

Table 3.2. Study 1 Pearson correlations of between-subjects variables and indices. * indicates < .05, ** indicates < .01, *** indicates > .001.

Manipulation Check. As a manipulation check of hindsight bias, we compared the average accuracy as a proportion (for participants in the hindsight condition, whether they reported they would have been accurate) using a 3 (training conditions) x 2 (training length) analysis of variance (ANOVA). As expected, there was a significant effect of condition ($F(2, 836) = 42.19$, $p < .001$, $\eta^2 = .09$), such that participants in the hindsight condition indicated they would have answered more trials correctly ($M = .86$, $SE = .01$) than participants actually answered in the foresight condition ($M = .75$, $SE = .01$, $t = 7.87$, $p_{Tukey} < .001$, $d = .68$) and foresight-with-feedback ($M = .75$, $SE = .01$, $t = 8.08$, $p_{Tukey} < .001$, $d = .68$) condition. There was no significant main effect of length ($F(1, 836) = 1.49$, $p = .22$, $\eta^2 = .002$), nor was there a significant interaction between condition and length, $F(2, 836) = 0.79$, $p = .45$, $\eta^2 = .002$. See Figure 3.1.

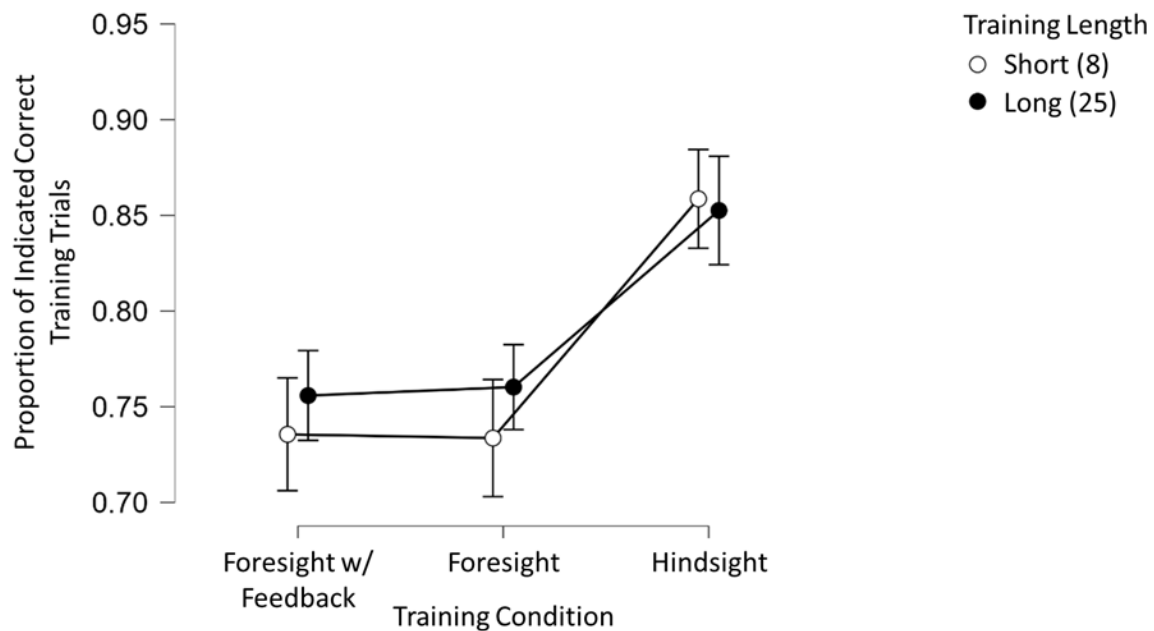


Figure 3.1. Average proportion of correct answers indicated during training trials by training condition and training length. Error bars represent 95% confidence intervals.

There are no significant effects of training condition ($F(2, 836) = 0.31, p = .73, \eta^2 < .001$) nor length ($F(1, 836) = 0.52, p = .47, \eta^2 < .001$) with respect to accuracy on the test trials. See Figure 3.2

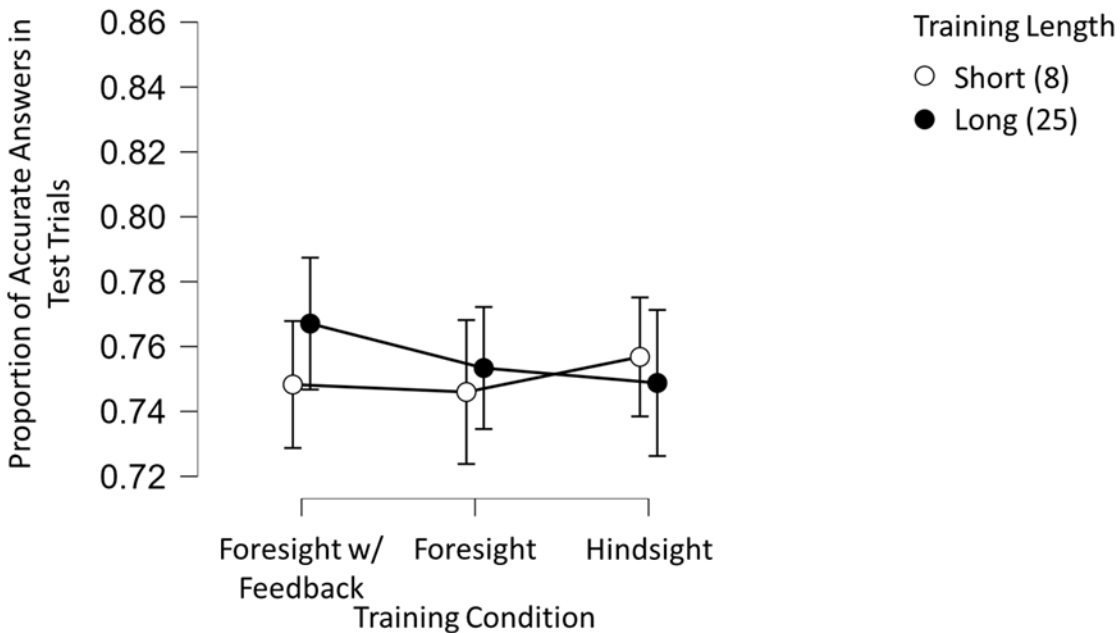


Figure 3.2. Average proportion of correct answers during test trials by training condition and training length. Error bars represent 95% confidence intervals.

Hindsight Effects on Prospective Confidence, Efficacy, and Difficulty Judgments were

largely similar. Using 3 (training conditions) x 2 (training length) ANOVAs, we evaluated condition differences in *prospective confidence* for the 45 trials, task efficacy beliefs, and difficulty beliefs for the task, thus testing H1.1 and H1.3, respectively, that repeated hindsight experience would increase prospective confidence for the test trials relative to foresight training and that more hindsight experiences would increase prospective confidence even more. First, the training conditions had a significant effect on prospective confidence in the test trials, $F(1, 836) = 19.07, p < .001, \eta^2 = .04$. Participants' prospective confidence was greater following training in hindsight ($M = .64, SE = .01, t = 4.66, p_{Tukey} < .001, d = .40$) and in foresight with feedback ($M =$

.66, $SE = .01$, $t = 5.86$, $p_{Tukey} < .001$, $d = .49$) than in foresight, $M = .56$, $SE = .01$. Prospective confidence did not significantly differ between participants who received training in hindsight and foresight with feedback, $t = -1.12$, $p_{Tukey} = .50$, $d = .10$; thus, H1.1 was supported. There was no significant main effect of length ($F(1, 836) = 2.02$, $p = .16$, $\eta^2 = .002$), nor was there a significant interaction between condition and length, $F(2, 836) = 2.33$, $p = .10$, $\eta^2 = .005$. Thus, H1.3 was not supported.

This pattern of results held when prospective confidence was replaced with *prospective confidence bias* (i.e., the difference between predicted and actual accuracy for the test phase). The training conditions had a significant effect on prospective confidence bias, $F(1, 836) = 16.50$, $p < .001$, $\eta^2 = .04$. Participants' prospective confidence bias was more positive following training in hindsight ($M = -0.11$, $SE = 0.01$, $t = 4.48$, $p_{Tukey} < .001$, $d = 0.38$) and in foresight with feedback ($M = -0.10$, $SE = 0.01$, $t = 5.37$, $p_{Tukey} < .001$, $d = .45$) than in foresight, $M = -0.18$, $SE = 0.01$. Prospective confidence bias did not significantly differ between participants who received training in hindsight and foresight with feedback, $t = -0.82$, $p_{Tukey} = .69$, $d = .07$. There was no significant main effect of length ($F(1, 836) = 3.54$, $p = .06$, $\eta^2 = .004$), nor was there a significant interaction between condition and length, $F(2, 836) = 2.73$, $p = .07$, $\eta^2 = .006$.

However, evaluating the *change in prospective confidence* (i.e., the difference between baseline prospective confidence and post-training prospective confidence) revealed more complex effects of the training manipulations. There was a significant main effect of training condition ($F(2, 836) = 20.42$, $p < .001$, $\eta^2 = .05$) and a significant interaction of training condition and length ($F(2, 836) = 4.22$, $p = .02$, $\eta^2 = .01$) on change in prospective confidence. There was no main effect of length, $F(1, 836) = 0.13$, $p = .72$, $\eta^2 < .001$. Tukey post-hoc analysis revealed that in the short conditions, no significant differences existed between the training conditions. However, in the

long conditions, there was a more positive change in prospective confidence for participants who received training in hindsight ($M = .02, SE = .01, t = 5.51, p_{Tukey} < .001, d = .68$) and in foresight with feedback ($M = .02, SE = .01, t = 5.60, p_{Tukey} < .001, d = .67$) than in foresight, $M = -.07, SE = .01$. However, change in prospective confidence did not significantly differ between participants who received training in hindsight and foresight with feedback in the long condition, $t = 0.11, p_{Tukey} = .99, d = -.01$. See Figure 3.3. Moreover, participants in the hindsight and foresight-with-feedback did not significant increase in prospective confidence, regardless of the length of training.

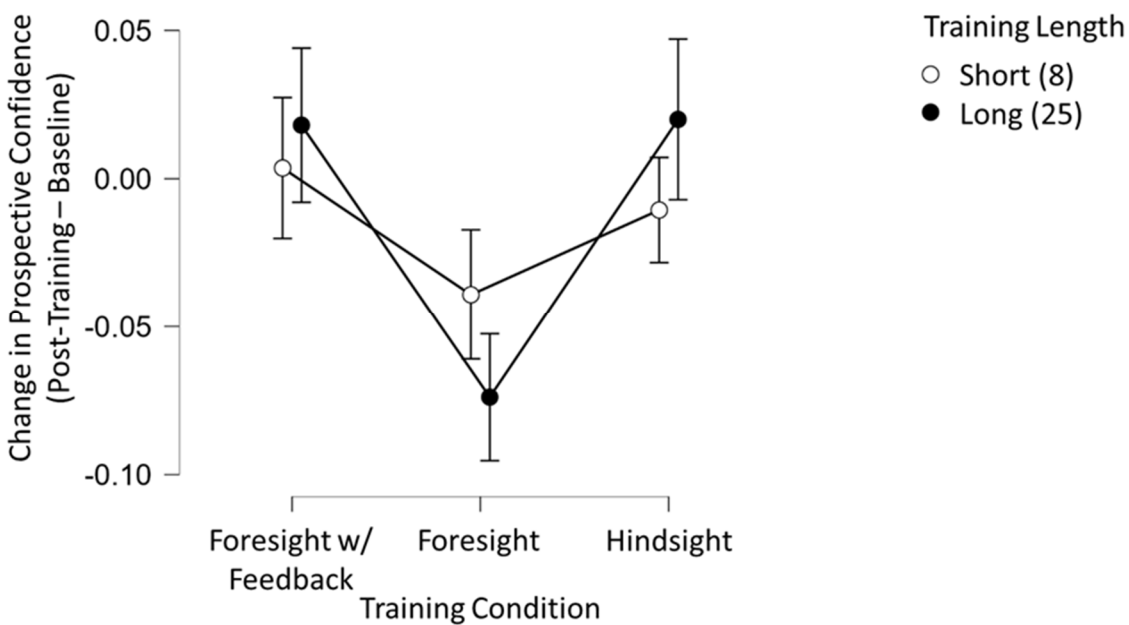


Figure 3.3. Average change in prospective confidence following training by training condition and training length. Error bars represent 95% confidence intervals.

The training condition also had a similar significant main effect on participants' *efficacy beliefs*, $F(2, 836) = 13.32, p < .001, \eta^2 = .03$. Participants reported that they were better at the task after training in hindsight ($M = 3.87, SE = 0.07, t = 3.05, p_{Tukey} = .007, d = .26$) and in foresight with feedback ($M = 4.06, SE = 0.06, t = 5.14, p_{Tukey} < .001, d = .43$) than in foresight ($M = 3.58, SE =$

0.07). Efficacy did not significantly differ between participants who received training in hindsight and foresight with feedback, $t = 2.04$, $p_{Tukey} = .10$, $d = .17$. There was no significant main effect of length ($F(1, 836) = 0.89$, $p = .35$, $\eta^2 = .001$), nor was there a significant interaction between condition and length, $F(2, 836) = 0.54$, $p = .58$, $\eta^2 = .001$.

The training condition, again, had a similar significant main effect on participants' *difficulty perceptions*, $F(2, 823) = 7.30$, $p < .001$, $\eta^2 = .02$. Participants reported that the task was less difficult after training in hindsight ($M = 4.15$, $SE = 0.06$, $t = -2.28$, $p_{Tukey} = .046$, $d = -.21$) and in foresight with feedback ($M = 4.06$, $SE = 0.06$, $t = -3.78$, $p_{Tukey} < .001$, $d = -.32$) than in foresight ($M = 4.35$, $SE = 0.06$). Perceived difficulty did not significantly differ between participants who received training in hindsight and foresight with feedback, $t = 1.35$, $p_{Tukey} = .37$, $d = .12$. There was no significant main effect of length ($F(1, 823) = 0.01$, $p = .93$, $\eta^2 < .001$), nor was there a significant interaction between condition and length, $F(2, 823) = 0.61$, $p = .54$, $\eta^2 = .001$.

There were no significant effects of the training manipulations on participants' self-reported *numeric self-efficacy*.

Hindsight Effects Also Were Similar for Retrospective Confidence. To test H1.2 and H1.4 (respectively, that repeated hindsight experience would increase retrospective confidence and that a greater number of hindsight experiences would increase retrospective confidence more), we fitted a 3 (Training Condition) x 2 (Length) ANOVA predicting overall retrospective confidence. There was a significant main effect of training condition on overall retrospective confidence, $F(2, 836) = 9.65$, $p < .001$, $\eta^2 = .02$. Participants who had undergone training in foresight with feedback ($M = .63$, $SE = 0.01$, $t = 4.36$, $p_{Tukey} < .001$, $d = .37$) and hindsight ($M = .61$, $SE = 0.01$, $t = 2.65$, $p_{Tukey} = .02$, $d = .23$) reported greater overall retrospective confidence in

their test performance than those in foresight, $M = .56$, $SE = 0.01$. There was no significant difference in overall retrospective confidence between participants in the hindsight and the foresight with feedback conditions, $t = -1.67$, $p_{Tukey} = .22$, $d = -.14$. There was no significant effect of the training length ($F(1, 836) = 1.42$, $p = .23$, $\eta^2 = .002$) nor an interaction between training condition and length, $F(2, 836) = 0.61$, $p = .55$, $\eta^2 = .001$. See Figure 3.4.

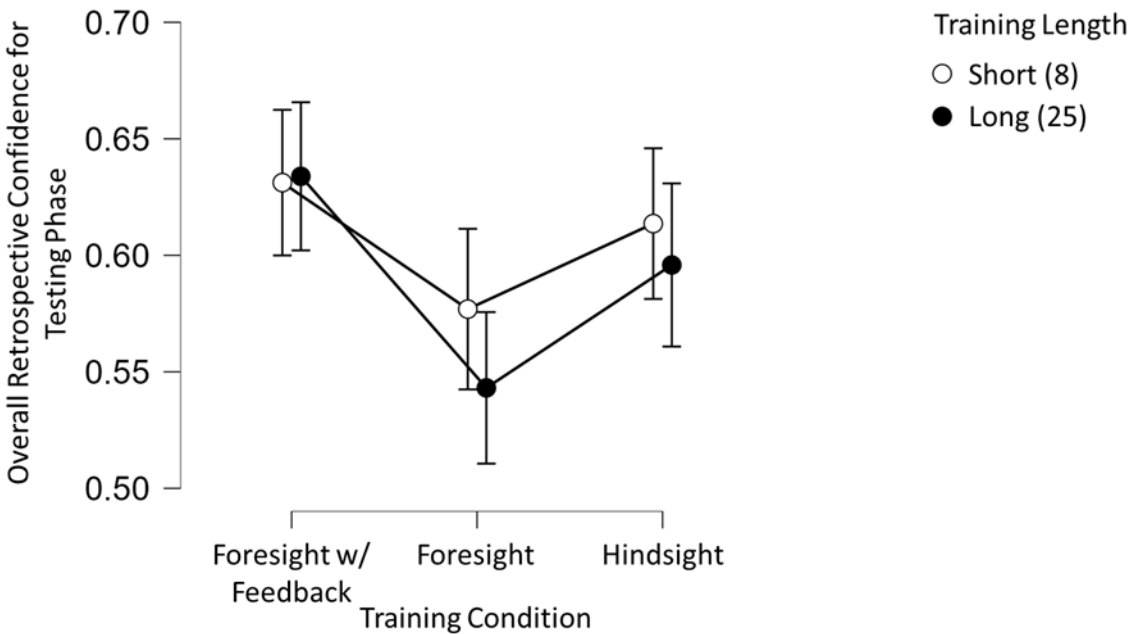


Figure 3.4. Average overall retrospective confidence judgments for the testing phase by training condition and training length. Error bars represent 95% confidence intervals.

Given the effects of the training condition on prospective and overall retrospective confidence, we sought to test whether the hindsight condition caused greater retrospective than foresight through its effect on prospective confidence. Thus, we fit a mediation model to investigate how the hindsight condition versus the foresight condition impacted participants' retrospective confidence through prospective confidence. There was a significant indirect effect of the training condition on retrospective confidence through prospective confidence ($ab = .06$, $z = 4.78$, $p < .001$) and no significant direct effect of condition on retrospective confidence ($c = -.01$, $z = 1.08$,

$p = .28$). See Figure 3.5.

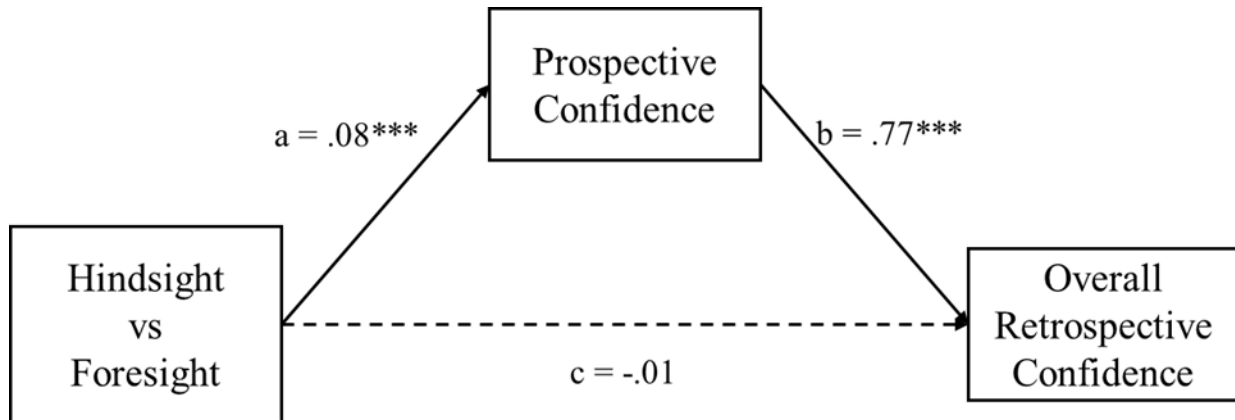


Figure 3.5. Mediation model depicting indirect effects of training condition (only the hindsight and foresight conditions) on overall retrospective confidence through prospective confidence. *** indicates $> .001$.

Next, we used a hierarchical linear regression with random intercepts for each participant and each trial to predict trial-by-trial retrospective confidence from the training condition, training length, and an interaction of training condition and length. These results differed from the previous retrospective confidence results. The model did not show any significant increase in trial-by-trial retrospective confidence in the foresight with feedback nor hindsight conditions over the foresight condition, $F(2,836.6) = 0.91, p = .40$. Moreover, there was no significant effect of training length ($F(1,836.6) = 0.02, p = .90$), nor any interactions of length and conditions, $F(2,836.6) = 2.32, p = .10$.

However, given that response times have a known effect on retrospective confidence (slower responses act as a heuristic to lower retrospective confidence; Baranski & Petrusic, 1998), an exploratory hierarchical regression model was fit predicting trial-by-trial retrospective confidence from the training condition, trial-by-trial response times, and the interaction of condition and response times, $Marginal R^2 = .03, Conditional R^2 = .50$. There was a significant

main effect of training condition ($F(2, 1640) = 8.28, p < .001$) and of response times, $F(1, 34966) = 1193.92, p < .001$. There was a significant interaction ($F(2, 34937) = 12.75, p < .001$) such that, when participants answered quickly, participants who experienced training in hindsight ($b = 0.04, SE = .02, p = .01$) and foresight with feedback ($b = 0.06, SE = .01, p < .001$) were more confident in their answers than participants in the foresight condition; the conditions did not differ when participants responded more slowly. Thus, there was greater negative effect of response times on trial-by-trial retrospective confidence for participants in the foresight with feedback ($\Delta b = 0.01, SE = 0.003, z = 5.04, p < .001$) and hindsight ($\Delta b = 0.01, SE = 0.003, z = 2.87, p = .01$) conditions than in the foresight condition. See Figure 3.6.

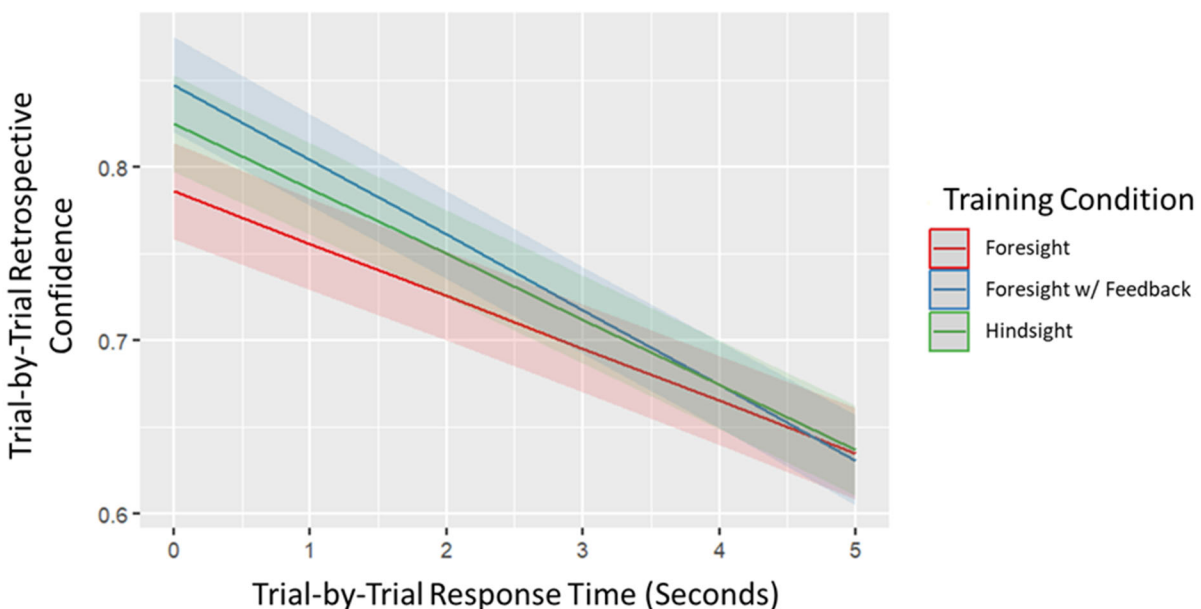


Figure 3.6. Estimated marginal means of trial-by-trial retrospective confidence by training condition and response time. Bands represent 95% confidence intervals.

Next, because we wanted to test whether the prospective confidence predicted retrospective confidence after accounting for accuracy (both at the trial level and participant level) and response times, we explored whether we could predict trial-by-trial retrospective confidence

judgments from the prospective confidence, trial-by-trial accuracy, trial-by-trial response times, the metacognitive awareness subscales, and overall accuracy as a proxy for ability, *Marginal R*² = .23, *Conditional R*² = .52. All independent variables other than accuracy were standardized. We found that greater prospective confidence predicted greater retrospective confidence ($F(1, 838) = 318.51, p < .001$), participants were more confident in fast answers than slow answers ($F(1, 35088) = 1006.63, p < .001$), participants were more confident when correct than when incorrect ($F(1, 34390) = 2710.92, p < .001$), and participants who reported a greater (vs lesser) tendency to monitor themselves were more confident, $F(1, 837) = 5.28, p = .02$. We also found a significant interaction between prospective confidence for the test and trial-by-trial response time ($b = -0.01, SE = .001, t = -6.99, p < .001$; see Figure 3.7A) such that prospective confidence was a stronger predictor of trial-by-trial retrospective confidence judgments for fast decisions than slow decisions. We also found a significant interaction between overall test accuracy and trial-by-trial accuracy ($b = 0.04, SE = .002, t = 15.38, p < .001$; see Figure 3.7B) such that participants who performed better on the test also demonstrated greater sensitivity of their retrospective confidence judgments to their trial-by-trial accuracy.

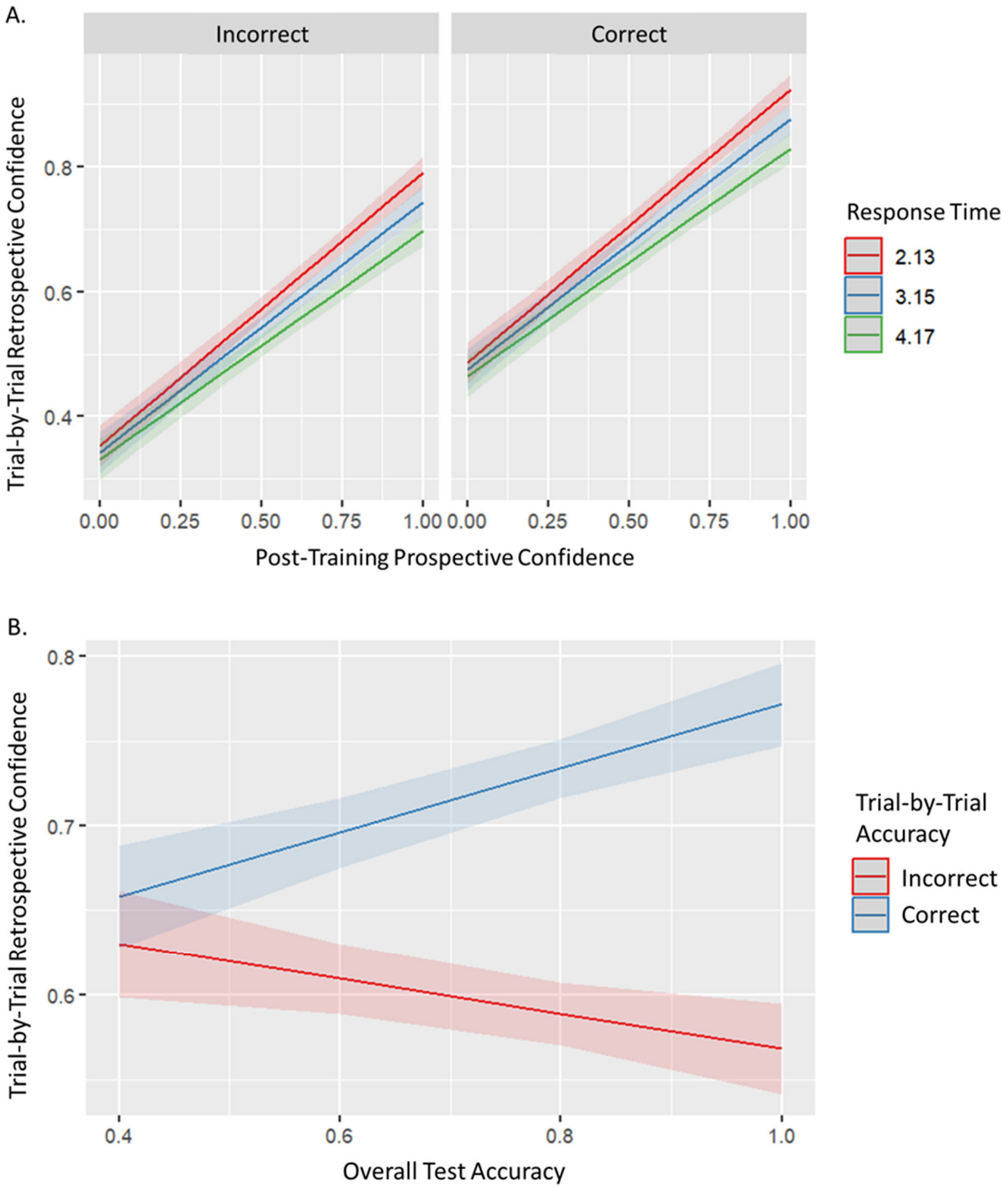


Figure 3.7. Estimated marginal means of trial-by-trial retrospective confidence by prospective confidence and response time (separate lines represent -1 standard deviation, the average, and $+1$ standard deviation; A) and by overall test accuracy and trial-level accuracy (B). Bands represent 95% confidence intervals.

Hindsight had no Effect on Retrospective Confidence Calibration and Sensitivity. We investigated RQ1.1 (whether an effect of hindsight experiences exists on retrospective confidence sensitivity to accuracy) by evaluating condition differences in both calibration (the retrospective Brier scores) and retrospective confidence sensitivity using separate 3 (training conditions) x 2 (training length) ANOVAs. There were no significant effects of the manipulation on calibration or sensitivity. There were no significant main effects of training conditions ($F(2, 836) = 1.53, p = .22$) nor training length ($F(1, 836) = 0.19, p = .66$) on participants' retrospective confidence Brier scores. There was also no significant interaction between condition and length, $F(2, 836) = 0.69, p = .50$. There were no significant main effects of training conditions ($F(2, 833) = 0.38, p = .68$) nor training length ($F(1, 833) = 0.01, p = .93$) on participants' retrospective confidence sensitivity. There was also no significant interaction between condition and length, $F(2, 833) = 0.40, p = .67$.

Metacognitive Awareness as a Moderator. To investigate RQ1.2 (that metacognitive awareness would moderate the effects of the training conditions on prospective and retrospective confidence), we fit a linear regression predicting prospective confidence bias (i.e., the difference between predicted performance and actual performance in the test phase) from the monitoring subscale of metacognitive awareness, the training conditions, the training length, and their interactions, $F(11, 830) = 6.20, p < .001, R^2 = .08$. There was a significant three-way interaction between monitoring, training condition, and length. Regardless of length, participants who reported greater monitoring tendencies were less underconfident on average for the test following training in hindsight, $b = 0.03, SE = 0.02, 95\% CI: [.0001, 0.06]$. Participants who reported greater monitoring tendencies were less underconfident for the test following training in foresight without feedback in the short condition ($b = 0.10, SE = 0.03, 95\% CI: [0.05, 0.16]$), but

more underconfident in the long condition, $b = -0.05$, $SE = 0.02$, $95\% CI: [-0.09, -0.002]$. See Figure 3.8. There was no effect of the evaluation subscale.

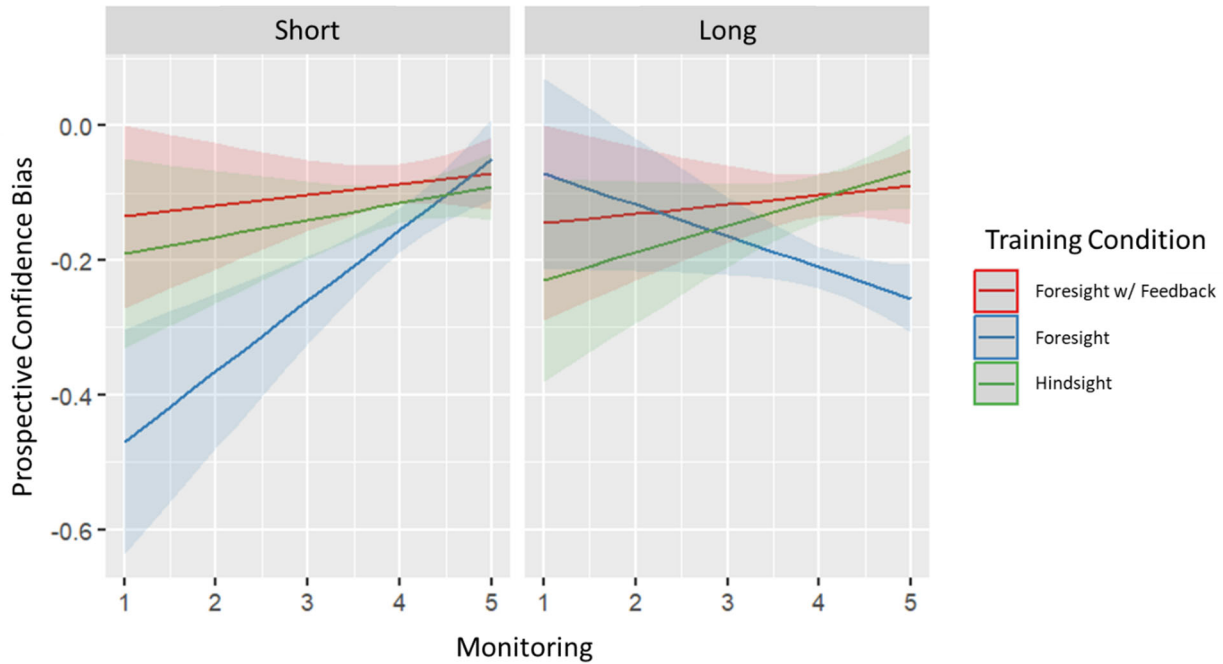


Figure 3.8. Estimated marginal means of prospective confidence bias for the test trials by condition, length, and monitoring. Bands represent 95% confidence intervals.

Similar results emerged for retrospective confidence bias. A linear regression was fit predicting retrospective confidence bias from the monitoring subscale of metacognitive awareness, the training conditions, the training length, and their interactions, $F(11, 830) = 3.54$, $p < .001$, $R^2 = .04$. There was a significant three-way interaction between monitoring, training condition, and length. Participants who reported greater monitoring tendencies were less underconfident on the test following training in foresight without feedback in the short condition ($b = 0.08$, $SE = 0.02$, $95\% CI: [0.04, 0.12]$), but not less underconfident in the long condition, $b = -0.01$, $SE = 0.02$, $95\% CI: [-0.04, 0.03]$. Participants who reported greater monitoring tendencies less underconfident on the test following training in foresight with feedback in the long condition (b

= 0.05, $SE = 0.02$, 95% $CI: [0.01, 0.09]$), but not less underconfident in the short condition, $b = -0.03$, $SE = 0.02$, 95% $CI: [-0.01, 0.06]$. Participants who reported greater monitoring tendencies were less underconfident on the test following training in hindsight in the long condition ($b = 0.06$, $SE = 0.02$, 95% $CI: [0.02, 0.10]$), but not less underconfident in the short condition, $b = -0.03$, $SE = 0.02$, 95% $CI: [-0.01, 0.06]$. See Figure 3.9. There was no effect of the evaluation subscale.

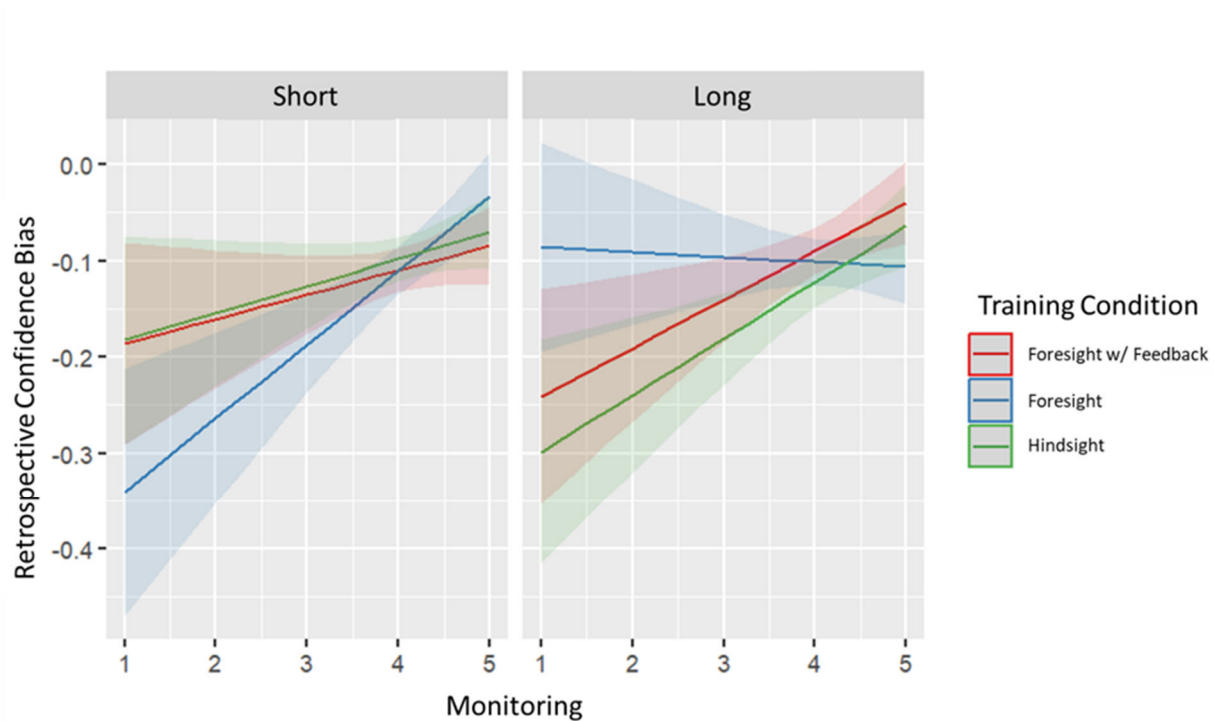


Figure 3.9. Estimated marginal means of retrospective confidence bias for the test trials by condition, length, and monitoring. Bands represent 95% confidence intervals.

Relation between Prospective Confidence and Deliberation. Finally, to test H1.5 (that higher prospective confidence after the training trials would motivate/demotivate deliberation on subsequent decisions in the test trials), we used a hierarchical linear regression with random intercepts for each participant and each trial to predict trial-by-trial response times from the *training condition*, training length, and overall test accuracy (standardized) as well as an

interaction of training condition and length, $Marginal R^2 = .01$, $Conditional R^2 = .33$. There was a marginal effect of the training condition on response times ($F(2, 834.94) = 2.65, p = .07$), no effect of length ($F(1, 834.94) = 2.21, p = .13$), and no interaction, $F(2, 834.94) = 1.11, p = .33$. However, participants who were overall more accurate on the test were faster at responding, $b = -0.05, SE = 0.02, t = -2.72, p = .007$. See Figure 3.10.

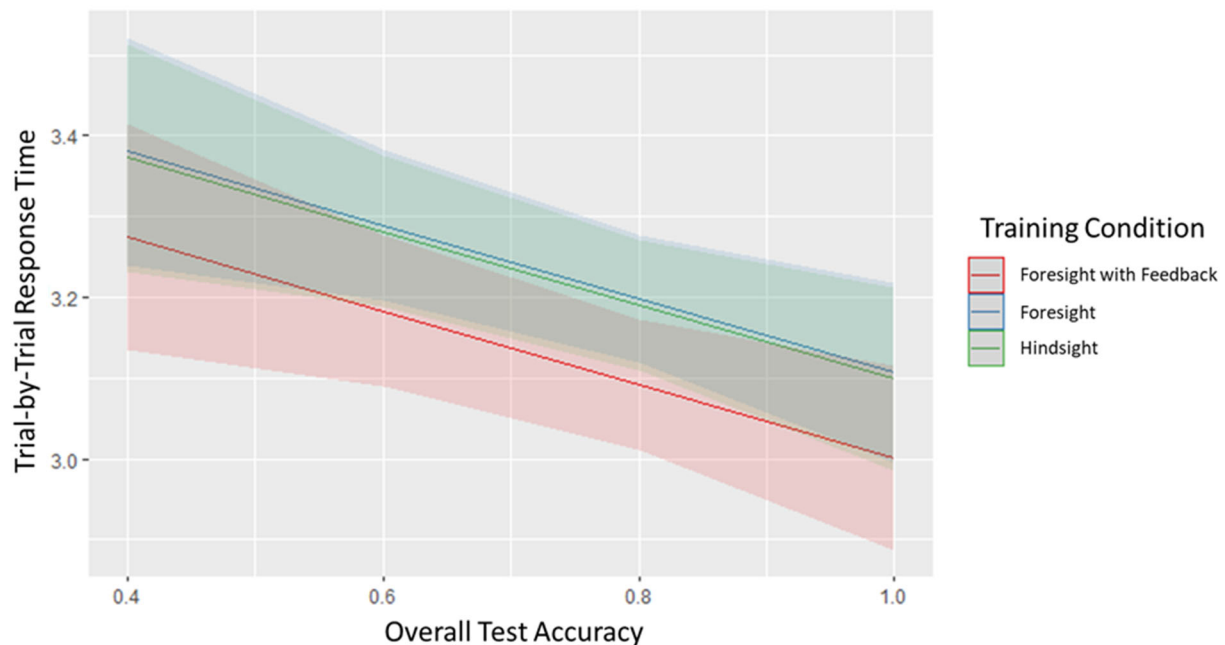


Figure 3.10. Estimated marginal means of response times for the test trials by condition and overall test accuracy. Bands represent 95% confidence intervals.

Next, we explored whether we could predict trial-by-trial response times from *prospective confidence*. To explore the nature of the relation between prospective confidence and response times, we fitted a linear model and a quadratic model. The quadratic model demonstrated better fit to the data than the linear model, $\chi^2 (df = 1) = 27.42, p < .001$; $Marginal R^2 = .02$, $Conditional R^2 = .33$. The quadratic model suggests that there is an inverse-U shaped relation between prospective confidence and response times, $b_{quadratic} = -2.14, SE = 0.41, t = -5.27, p < .001$; b_{linear}

= 2.81, $SE = 0.48$, $t = 5.86$, $p < .001$. See Figure 3.11.

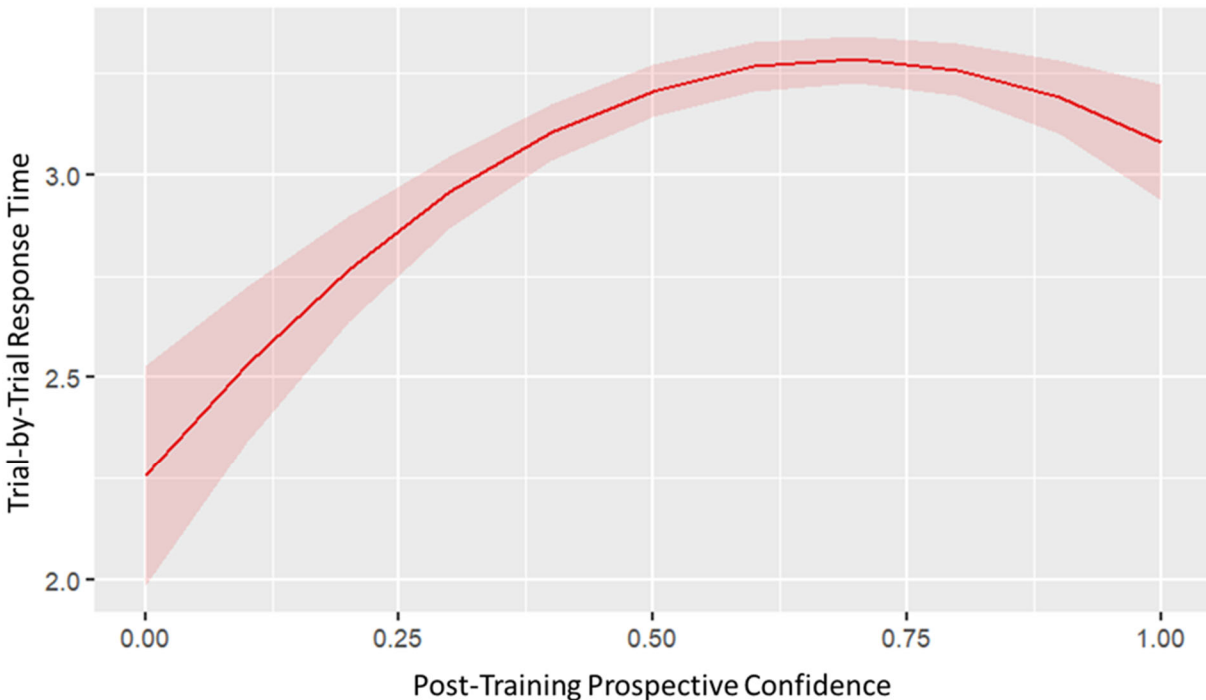


Figure 3.11. Estimated marginal means of response times for the test trials by prospective confidence. Bands represent 95% confidence intervals.

Next, we explored whether the quadratic relation of prospective confidence would persist when controlling for participants' accuracy. We predicted trial-by-trial response times from prospective confidence (quadratic and linear terms), trial-by-trial accuracy, and overall accuracy (standardized) as a proxy for ability and the interaction of trial-level accuracy and overall accuracy, $Marginal R^2 = .03$, $Conditional R^2 = .34$. In addition to the quadratic and linear effects of prospective confidence ($b_{quadratic} = -2.17$, $SE = 0.40$, $t = -5.41$, $p < .001$; $b_{linear} = 3.01$, $SE = 0.48$, $t = 6.02$, $p < .001$), participants were faster at answering when correct, $F(1, 33150) = 266.97$, $p < .001$. See Figure 3.12A. There was also a significant interaction between trial-level accuracy and overall test accuracy ($b = -0.14$, $SE = 0.01$, $t = -12.21$, $p < .001$) such that

participants who were overall more accurate were faster when they were correct than participants who were overall less accurate ($b = -0.90$, $SE = 0.17$, $95\% CI: [-1.24, -0.56]$) but they were not significantly different in speed when incorrect, $b = 0.30$, $SE = 0.19$, $95\% CI: [-0.07, 0.67]$. See Figure 3.12B.

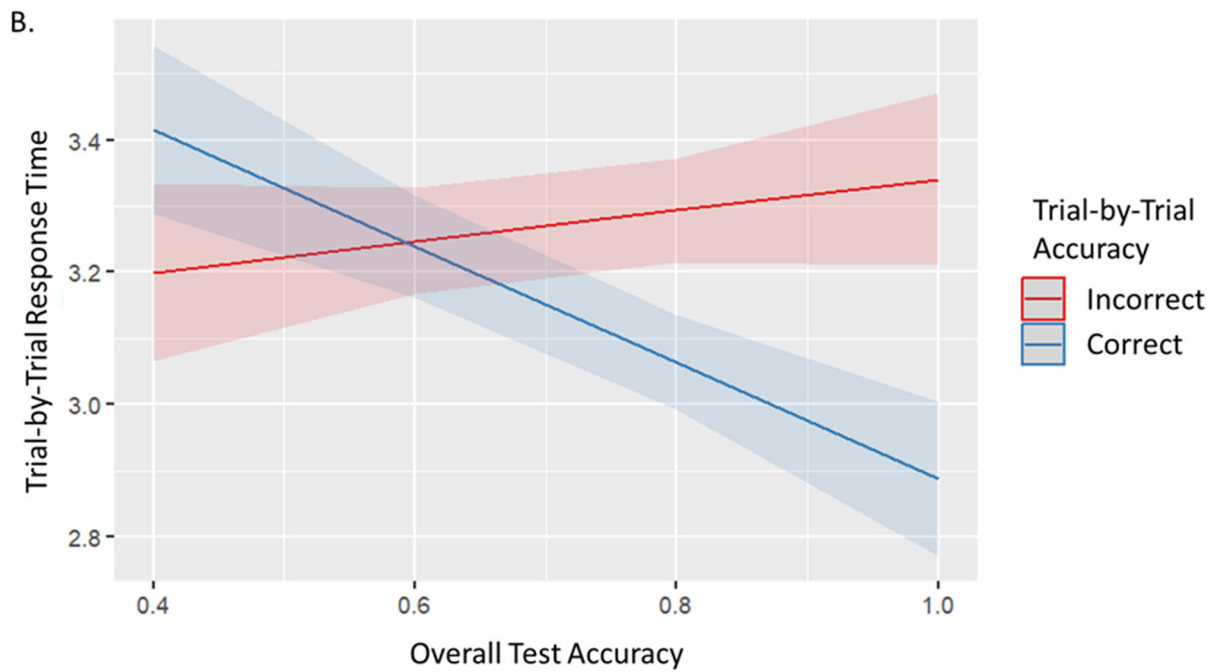


Figure 3.12. Estimated marginal means of trial-by-trial response times by prospective confidence and trial-level accuracy (A) and by overall test accuracy and trial-level accuracy (B). Bands represent 95% confidence intervals.

Affect to Outcomes. In preparation for Study 2, we explored whether there were condition differences in participants' reported affective responses by using four separate 2 (hindsight v. foresight with feedback) x 2 (training length) ANOVAs predicting valence toward correct, arousal toward correct, valence toward incorrect, and arousal toward perceived correct responses in training trials. We only evaluated two conditions because we were interested in affective responses to the outcome information (learn that they were/would have been correct/incorrect). This outcome information was not provided in the foresight condition. We found no significant effects of condition nor length for any of the affect measures.

For participants' reported affect valence to a perceived correct response, there was no significant main effect of condition ($F(1, 565) = 0.87, p = .35, \eta^2 = .002$), main effect of length ($F(1, 565) = 0.55, p = .46, \eta^2 < .001$), nor interaction, $F(1, 565) = 0.04, p = .85, \eta^2 < .001$. For participants' reported affective arousal to a perceived correct response, there was no significant main effect of condition ($F(1, 565) = 0.14, p = .71, \eta^2 < .001$), main effect of length ($F(1, 565) = 0.90, p = .34, \eta^2 = .002$), or interaction, $F(1, 565) = 1.71, p = .19, \eta^2 = .003$. For participants' reported affect valence to a perceived incorrect response, there was no significant main effect of condition ($F(1, 565) = 0.01, p = .93, \eta^2 < .001$), main effect of length ($F(1, 565) = 3.81, p = .052, \eta^2 = .01$), or interaction, $F(1, 565) = 0.13, p = .72, \eta^2 < .001$. Finally, there was no significant main effect of condition ($F(1, 565) = 1.41, p = .24, \eta^2 = .002$), main effect of length ($F(1, 565) = 0.39, p = .53, \eta^2 < .001$), or interaction on participants' reported affective arousal to a perceived incorrect response, $F(1, 565) = 1.10, p = .29, \eta^2 = .002$.

One-Week Follow-Up. For the one-week follow-up, we explored condition differences in prospective confidence for 45 hypothetical trials, task efficacy beliefs, and difficulty beliefs using three separate 3 (training conditions) x 2 (training length) ANOVAs. Thus, we tested the

durability of H1.1 and H1.3. There was a significant main effect of training condition on participants' prospective confidence after one week, $F(2, 755) = 10.54, p < .001, \eta^2 = .03$. Participants who trained in the foresight-with-feedback ($M = 0.66, SE = .01$) condition were more confident in their performance on a hypothetical test than those who trained in hindsight ($M = 0.60, SE = .01, t = 2.91, p_{Tukey} = .01, d = .26$) or in foresight, $M = 0.56, SE = .01, t = 4.52, p_{Tukey} < .001, d = .40$. Participants in the foresight and hindsight conditions did not significantly differ in their prospective confidence after one week, $t = -1.59, p_{Tukey} = .25, d = -.14$. There was no main effect of length ($F(1, 755) = 0.83, p = .36, \eta^2 = .001$), nor a significant interaction, $F(2, 755) = 0.79, p = .46, \eta^2 = .002$. See Figure 3.13.

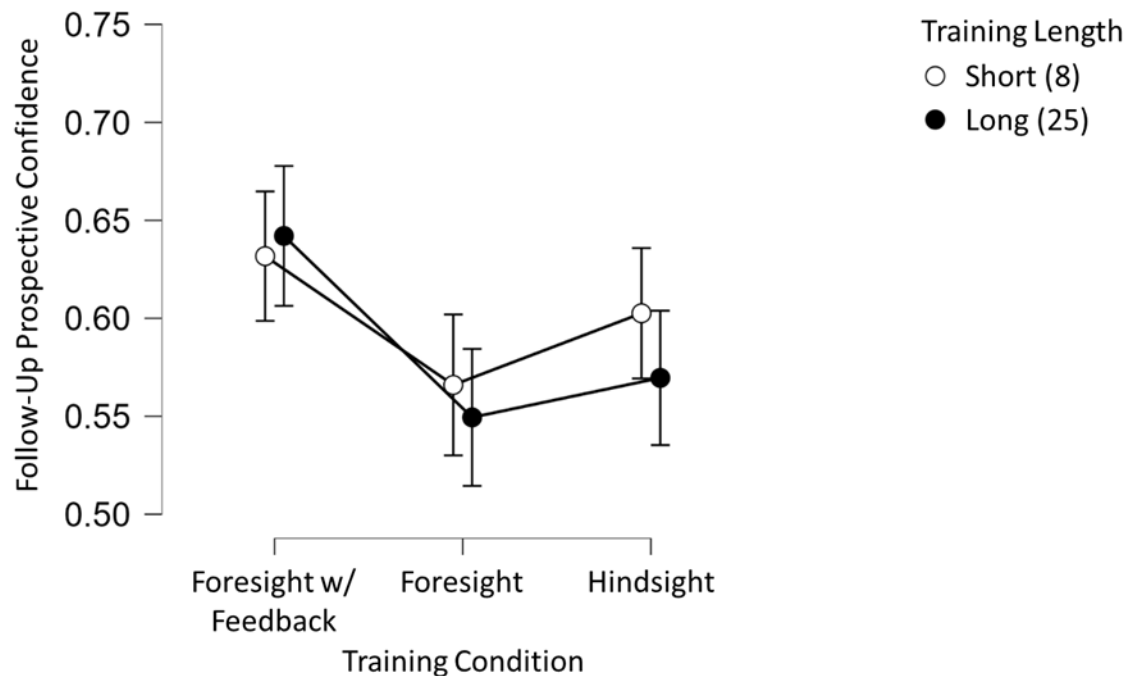


Figure 3.13. Average prospective confidence judgments during a one-week follow-up for a hypothetical test by training condition and training length. Error bars represent 95% confidence intervals.

Other results were substantially similar. There was a significant main effect of training condition on participants' task efficacy after one week, $F(2, 755) = 4.04, p = .02, \eta^2 = .01$. Participants who

trained in the foresight-with-feedback ($M = 3.95$, $SE = .07$) condition were more self-efficacious than those who trained in hindsight ($M = 3.72$, $SE = .07$, $t = 2.41$, $p_{Tukey} = .04$, $d = .21$) and in foresight, $M = 3.71$, $SE = .07$, $t = 2.48$, $p_{Tukey} = .04$, $d = .22$. Participants in the foresight and hindsight conditions did not significantly differ in their task efficacy after one week, $t = -0.08$, $p_{Tukey} = .997$, $d = -.01$. There was no main effect of length ($F(1, 755) = 0.04$, $p = .84$, $\eta^2 < .001$), nor a significant interaction, $F(2, 755) = 0.07$, $p = .93$, $\eta^2 < .001$.

Lastly, there was no significant main effect of training condition on participants' task difficulty judgments after one week ($F(2, 755) = 1.61$, $p = .20$, $\eta^2 = .004$), no main effect of length ($F(1, 755) = 0.55$, $p = .46$, $\eta^2 < .001$), nor a significant interaction, $F(2, 755) = 1.86$, $p = .16$, $\eta^2 = .01$.

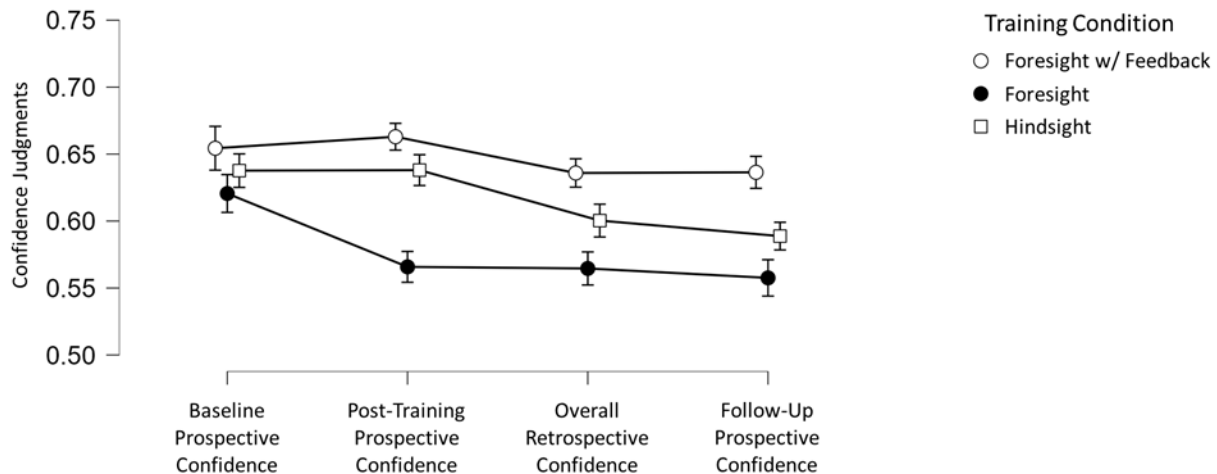


Figure 3.14. Average confidence judgments throughout the study by training condition. Error bars represent 95% confidence intervals. Post hoc used a Holm correction for a family of 66. Conditions did not significantly differ at baseline. Post-training, prospective confidence was significantly lower for foresight (hindsight and foresight with feedback did not significantly differ). For overall retrospective, foresight and foresight with feedback significantly differed from each other, but neither significantly differed from hindsight. At the follow-up, again, foresight and foresight with feedback significantly differed from each other, but neither significantly differed from hindsight.

We employed a repeated measures ANOVA evaluating differences in confidence judgments by condition and by when the judgment was made. Judgments included were the baseline prospective confidence, the post-training prospective confidence, the overall retrospective confidence, and the follow-up prospective confidence judgments. There was a significant main effect of time ($F(3, 2274) = 30.40, p < .001, \eta^2 = .01$) such that subsequent judgments tended to be lower than the immediate prior judgment other than the follow-up prospective confidence which did not significantly differ from overall retrospective confidence, $\Delta M = 0.01, SE = 0.004, t = 1.41, p_{Holm} = .16, d = .03$. Additionally, there was a significant main effect of training condition on confidence judgments, $F(2, 758) = 10.34, p < .001, \eta^2 = .02$. Participants in the foresight-with-feedback condition were overall more confident than participants in hindsight ($\Delta M = 0.03, SE = 0.02, t = 2.04, p_{Holm} = .04, d = .16$) and in foresight, $\Delta M = 0.07, SE = 0.02, t = 4.55, p_{Holm} < .001, d = .36$. Participants in hindsight condition were overall more confident than participants in foresight, $\Delta M = 0.04, SE = 0.02, t = 2.50, p_{Holm} = .03, d = .20$. Finally, there is a significant interaction between time and training condition ($F(6, 2274) = 6.03, p < .001, \eta^2 = .003$), such that participants' confidence judgments following training in the foresight-with-feedback condition appeared to maintain their more elevated level of confidence more than participants in the hindsight condition. See Figure 3.14.

Discussion

As expected, participants in the hindsight condition indicated that they would have answered more items correctly than participants in the foresight and foresight with feedback conditions got correct, suggesting a successful hindsight manipulation (Hoch & Loewenstein, 1989). Further, we expected and showed no proportionate difference in accuracy for the training choices between participants in the long and short training. After the training trials, those in the hindsight

condition indicated greater prospective confidence for a subsequent test than participants in the foresight condition, thus H1.1 was supported. However, our results do not support the speculation that hindsight bias *increases* confidence. Rather, in the present study, participants in the foresight condition indicated a reduced confidence following the training phase, while participants in the hindsight and foresight-with-feedback conditions maintained their level of confidence. There were mixed results of the effect of the training length on prospective confidence (H1.3). The hypothesized interaction of training length and condition did emerge when assessing participants' change in their prospective confidence from baseline; however, the pattern of results was more consistent with a greater number of trials in the foresight condition lowering prospective confidence rather than more trials in the hindsight condition raising prospective confidence. Thus, H1.3 was generally not supported. The decrease in confidence in the foresight condition and lack of change in the hindsight condition may be because participants were on average underconfident for the choice task (average prospective confidence bias was -0.13). Additionally, the five-second time limit might have caused participants in the foresight condition to have low retrospective confidence for the training trials. Indeed, interrogation paradigm studies (when participants are stopped mid-decision) suggest that when people are forced to make faster decisions, their retrospective confidence is lower (Vickers et al., 1985). Participants in the hindsight and foresight-with-feedback condition may have maintained their confidence because of the hindsight bias and objective performance information, respectively.

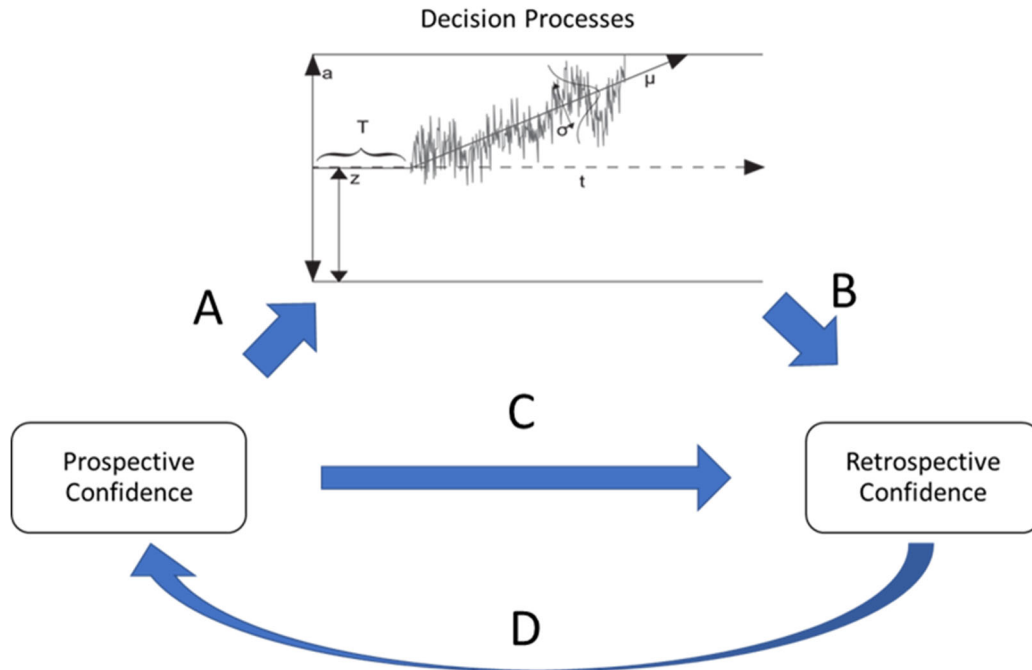


Figure 3.15. *The Confidence Cycle. Path A represents the function of prospective confidence as a motivator/demotivator of deliberation. Path B represents the influence of the decision-making experience (decision processes are represented in line with the Drift-Diffusion Model) on retrospective confidence. Path C represents the function of prospective confidence in informing retrospective confidence directly. Lastly, Path D represents the function of retrospective confidence in updating prospective confidence.*

There is also support for H1.2—that training in hindsight would increase retrospective confidence in subsequent decisions. However, there were no significant effects of training length on retrospective confidence and thus no support for H1.4. The effect of training condition on retrospective confidence judgments (both overall and trial-by-trial) supports a causal relation between prospective confidence and retrospective confidence (Path C in Figure 3.15). However, it remains plausible that a factor other than prospective confidence explains the effects of the training condition on retrospective confidence. Further, prospective confidence bias is positively associated with retrospective confidence bias ($r = .49, p < .001$), suggesting that, consistent with prior literature (i.e., Ehrlinger & Dunning, 2003; Van Marcke et al., 2024), prospective confidence is an important determinant of retrospective confidence. However, we find that

prospective confidence is not the only predictor of trial-by-trial retrospective confidence. Consistent with prior findings, participants were less confident retrospectively when they were slower (Baranski & Petrusic, 1998). Response time or even experienced fluency may have acted as a cue informing participants that the trial was difficult for them (Alter & Oppenheimer, 2009). Interestingly, response times appeared to water down the influence of prospective confidence on retrospective confidence. Specifically, it may be that when people respond more slowly, their response time and their experienced fluency become a more prominent indicator of their accuracy relative to their expectations and prospective confidence. Additionally, we find that the retrospective confidence judgments of participants who performed better on the test overall were more sensitive to their accuracy. The positive association between performance and retrospective confidence sensitivity is consistent with the idea that confidence and choice are based on the same evidence accumulation processes and skills (Kruger & Dunning, 1999; Pleskac & Busemeyer, 2010). These findings regarding response times and retrospective confidence support Path B in the Confidence Cycle (see Figure 3.15). Retrospective confidence appears to be directly informed by prospective confidence but is also sensitive to the specific decision experiences.

Somewhat consistent with H1.5—that prospective confidence would increase deliberation—prospective confidence had a parabolic relation with response times. This finding may reconcile previous conflicting findings that greater prospective confidence increases (Boldt et al., 2019; van den Berg et al., 2016; Peters & Shoots-Reinhard, 2022; Zacharopoulos et al., 2014) and decreases deliberation (Alter et al., 2007; Dutilh et al., 2012; Vancouver et al., 2002; Bandura & Jourden, 1991). However, we do not have causal evidence (the effect of condition was not significant, though marginal) for the relation between prospective confidence and deliberation.

The lack of causal evidence may be for a couple of reasons. First, given the non-linear relation between prospective confidence and response times, participants may exhibit a positive or negative relation depending on their initial level of confidence. Sampling from participants varying in prospective confidence, therefore, may cancel out the effects of the manipulation. Second, response times are not a clean indicator of decision processes. Longer responses may be caused by participants requiring more evidence to make a decision or a slower accumulation of evidence. To account for the rate of evidence accumulation, we attempted to control for participants' ability at the task using overall test accuracy and its interaction with trial-by-trial accuracy. In the next chapter, we fit a drift diffusion model that will allow us to estimate parameters for decision thresholds and drift rates to further investigate the effects of prospective confidence on decision processes.

At the one-week follow-up, we still observed an effect of the training condition on prospective confidence; however, while confidence among participants in the foresight-with-feedback condition remained somewhat stable, participants in the hindsight condition reported lower prospective confidence than after the training phase. Across conditions, participants' prospective confidence after one week was more similar to their overall retrospective confidence judgments after the test phase than their prospective confidence following training, supporting the theorized updating function of retrospective confidence (Path D in Figure 3.15). In other words, participants updated their prospective confidence during the test phase, and that prospective confidence remained mostly stable after one week. However, the question remains: why did participants in the hindsight condition lose confidence during the test phase while participants in the foresight-with-feedback didn't lose confidence despite similar accuracy for the test trials? One possibility is that participants in the hindsight condition were surprised by the difficulty of

the test trials. This possibility seems unlikely given that participants in the hindsight condition reported relatively high trial-by-trial retrospective confidence in the test trials, especially for fast choices. Another possibility is that something about the experience of learning confidence in hindsight made it less durable. Although self-reported affect did not differ between conditions in the present Study 1, in Study 2, we examine the possibility that people's affective responses to perceived successes and failures are different in hindsight than in foresight with feedback.

The lack of a consistent length effect (H1.3 and H1.4) may suggest that participants considered the training phase as a whole rather than as multiple separate trials. It is also possible that participants only considered a subset of trials and neglected the overall durations (Kahneman et al., 1993). Given the lack of a length effect on the proportion of correct answers indicated, considering the whole training might explain why there was little effect of the length manipulation. This explanation is somewhat supported by the findings involving monitoring tendencies. Participants who reported high monitoring tendencies became more underconfident prospectively following the long foresight training compared to the short foresight training. It is possible that participants higher in monitoring were more likely to consider the training trials separately rather than as a whole and were thus more influenced by the length differences.

Although we found no condition differences in participants' retrospective confidence calibration and sensitivity, we see that greater prospective confidence is associated with retrospective confidence calibration as operationalized by Brier scores and greater sensitivity to accuracy. Thus, we do not find that hindsight bias reduces sensitivity to accuracy (as suggested by Dawson et al., 1993).

Limitations

There are several limitations to the present study. First, the study was collected online through MTurk and was completed by participants in uncontrolled environments. Participants may have been distracted by events in their environment, which could have impacted their performance and their response times. While we cannot rule out that distractions could have influenced correlational findings, it does not challenge the validity of the causal findings of the manipulations because participants were randomly assigned to conditions. Nonetheless, distractions may have reduced the power of the present study to find effects. As a result, in study 2, participants completed the study in-person in a relatively quiet and distraction-free environment.

Another limitation is that prospective confidence was measured for each phase (i.e., training and testing) rather than for each trial (as was done by Fleming et al., 2016). As such, the bias of participants' prospective confidence can be assessed, but sensitivity and Brier scores cannot. Moreover, we cannot assess the changes in prospective confidence throughout the training and testing phases in responses to participants' trial-level retrospective confidence. Moreover, retrospective confidence was not measured for the training phase. Had retrospective confidence been measured for the training phase, we could have explored the relationship between participants' retrospective confidence for the training and their prospective confidence for the test. We expect that these confidence judgments would be very similar—as was observed for participants' retrospective confidence for the test and prospective confidence in the one-week follow-up. Additionally, with more confidence judgments, we could more effectively investigate the trends and relations of prospective confidence, response times, and retrospective confidence longitudinally.

Affective responses were only assessed at the end of the initial collection (and not during the 1-week follow-up), and participants were asked to recall their feelings from the training trials.

Participants would likely have difficulty recalling their feelings accurately, and it would make it difficult to investigate differences in affective responses to feedback. In Study 2, participants were asked to indicate their affective responses in real-time following each training trial.

Participants were, by and large, underconfident when it came to this task. Participants may have lacked confidence because the task involved a lot of math and due to the 5-second time limit.

Despite participants' low confidence, we still observed a hindsight bias. The confidence-hindsight mirror effect would predict that participants who are underconfident might actually become less confident in hindsight (Winman et al., 1998); however, we did not observe a reduction of confidence in the hindsight condition. Nevertheless, participants' low confidence relative to their performance throughout might have reduced the magnitude of the hindsight bias. Had participants been overconfident, we might have observed larger effects of experience in hindsight on prospective confidence, response times, and retrospective confidence.

Lastly, we explicitly elicited prospective confidence judgments; this explicit elicitation might have increased their impact on response times and retrospective confidence because their retrospective confidence was made more salient. Future research could investigate how likely participants are to spontaneously judge their confidence prospectively for tasks and avoid assessing prospective confidence close to the task itself.

Significance and Contributions to the Field

First, our research can speak somewhat to how repeated hindsight bias can affect future decisions and confidence. Although there has been much speculation regarding the consequences of

hindsight bias (e.g., Dawson et al., 1993; Roese & Vohs, 2012), its effect on subsequent decision making, prospective, and retrospective confidence has not been causally shown. We directly tested the consequences of repeated hindsight experience for prospective confidence, retrospective confidence, and decision making. Relative to the foresight condition, affording participants hindsight bias in training caused greater prospective confidence and efficacy beliefs for the task—although it did not increase them. Moreover, participants’ greater prospective confidence also increased their retrospective confidence on subsequent trials.

Additionally, our findings help shed light on how prospective confidence modulates decision-making processes. Specifically, our findings suggest a nonlinear, parabolic relation between prospective confidence and deliberation. This finding may help reconcile opposing findings that suggest that prospective confidence could motivate (Boldt et al., 2019; van den Berg et al., 2016; Peters & Shoots-Reinhard, 2022; Zacharopoulos et al., 2014) and demotivate deliberation (Alter et al., 2007; Dutilh et al., 2012; Vancouver et al., 2002; Bandura & Jourden, 1991). In other words, our findings demonstrate that prospective confidence can motivate and demotivate deliberation, like classic achievement motivation theories (Atkinson, 1957). Additionally, the parabolic relation suggests that prior findings are dependent on participants’ overall level of prospective confidence and could be reversed. For example, Dutilh and colleagues (2012) found that participants become more cautious (i.e., make slower decisions) following a perceived error—post-error slowing. However, given the parabolic relation, post-error slowing may only be observed in participants who have very high prospective confidence overall. Participants with low prospective confidence may instead show “post-error quickening” if the perceived error further reduces their prospective confidence for the decision task.

The present study provides support for the framework of the Confidence Cycle described in Chapter 1. Prospective confidence appears to modulate decision processes (i.e., the amount of and the rate of evidence accumulation). Retrospective confidence is informed by the decision-making experience and by prospective confidence directly. Then, retrospective confidence updates prospective confidence. The Confidence Cycle can help explain how confidence changes over time and how the effects of confidence on decision making change dynamically.

Confidence appears to play an important role in determining how much people deliberate for a given decision, as well as assessing the outcome of that decision and learning from the decision-making experience. It is plausible that confidence serves to optimize people's effort for decisions by avoiding too much effort on decisions that are too hard, saving effort when facing easy decisions, and allocating effort for challenging decisions for which we are uncertain of the outcome (i.e., whether we will be correct or incorrect). Retrospective confidence allows us to learn when we should adjust our efforts for future similar decisions.

IV: STUDY 2: AFFECT'S ROLE IN LEARNING CONFIDENCE

Study 2 aimed to investigate the role of affect (i.e., positive and negative evaluations experienced as feelings) in learning confidence from known outcomes (either when getting feedback about a decision made in foresight or when seeing the correct answer in hindsight). The foresight condition from Study 1 was not included because we were interested in participants' affective responses to whether they were/would have been correct/incorrect. Participants never learned the answer in the foresight condition. The effect of integral affect toward decision outcomes on confidence has not been studied. Previous research suggests that incidental affect influences confidence judgments. For example, people with depression have reported lower retrospective confidence judgments even after accounting for performance (Fieker et al., 2016; Fu et al., 2005). Consistent with these results, Culot and colleagues (2021) showed that inducing incidental sadness reduced retrospective confidence judgments for a perceptual decision task without an effect on performance. Similarly, positive mood inductions have been shown to increase retrospective overconfidence (Ifcher & Zarchamee, 2014; Koellinger & Treffers, 2015). Positive moods have also been associated with greater self-efficacy (see Lyubomirsky et al., 2005), as well as increased prospective confidence for laboratory tasks (Wright & Mischel, 1982) and increased attributions for successes to stable, internal factors (Brown, 1984). While these studies together demonstrate that affective dispositions and states influence prospective and retrospective confidence judgments, they only speak to the effects of incidental affect.

The focus of the present study concerns the role of integral affect toward being correct or incorrect in learning prospective confidence and informing retrospective confidence. We speculated that affective responses to decision outcomes in training would inform participants' prospective confidence judgments similar to the somatic marker hypothesis (Bechara et al.,

1997). Bechara and colleagues demonstrated that experienced anticipatory affect (operationalized as participants' electrodermal activity) was not only important for participants' ability to learn the outcomes associated with a given action but also to make advantageous decisions. Similar to their findings, we suggest that affective responses to being correct and incorrect will help people learn how good/bad they are at a decision task. Further, we sought to investigate how learning in hindsight might impede people's ability to learn to make accurate and sensitive confidence judgments by changing their affective responses while gaining experience. We suspected that in hindsight, participants would perceive they would have been correct more and experience positive affect more often, regardless of whether they would have actually been correct. As such, their anticipatory affect won't help them differentiate between questions they would and would not answer correctly.

In Study 2, integral affect was measured in two ways: 1) self-report of experienced affective valence and arousal, and 2) measurement of participants' electrodermal activity (EDA). However, the present dissertation will focus solely on participants' self-reported affect. This self-report measure has two advantages over measuring EDA: 1) it is a less noisy and more direct indicator of anticipatory affect, and 2) it can capture variance in participants' affective valence. However, it might be more subject to demand effects. Mellers and colleagues (2013) speculated that affect responses are diminished in hindsight compared to foresight because people experience less surprise in hindsight. Muted affect responses in hindsight may be supported by the Study 1 finding that participants' confidence persisted longer following training in the foresight-with-feedback condition than in the hindsight condition given that past research has shown that greater affect responses support long-term memory (Kensinger, 2009; Mather, 2007; Peters, Shoots-Reinhard, et al., 2019). For the present study, we hypothesized that:

H2.1) During training, participants would feel more negative after *negative feedback* (i.e., when a participant either learns they were incorrect or perceives that they would have been incorrect) compared to *positive feedback* (i.e., when a participant either learns they were correct or perceives that they would have been correct);

H2.2) H2.1's affective response sensitivity would be moderated by training condition (foresight with feedback v. hindsight; H2.2a) and by metacognitive awareness (H2.2b) such that affective responses would be more sensitive to accuracy in foresight with feedback compared to hindsight and for participants higher in metacognitive awareness, respectively;

H2.3) Valence and arousal of affective responses would predict changes in prospective confidence, such that positive responses would increase prospective confidence and negative responses would decrease prospective confidence, with higher arousal responses having a greater impact;

RQ2.1) People higher vs lower in metacognitive awareness would update their prospective confidence more in response to their emotional responses (RQ2.1).

Study 2 also seeks to replicate and extend key findings from Study 1 regarding the relations between prospective confidence, decision processes, and retrospective confidence. To investigate how prospective confidence influences decision making, we will employ drift diffusion modeling (DDM; Ratcliff & Rouder, 2000). The DDM can be thought of as an extension of Signal Detection Theory. The DDM provides a process model—a random accumulation of evidence over time—to explain the observed choices and response times in a two-alternative forced-choice task. In the DDM, evidence is accumulated until a decision threshold is met, at which point a

decision is made. Based on participants' multiple decisions and response times, the model estimates several parameters (Ratcliff, 1978; see Figure 4.1):

1. Drift rate (v): The rate at which evidence accumulates for one option relative to another.
2. Decision threshold (a): The amount of evidence required to make a decision. This parameter is sometimes referred to as decision cautiousness.
3. Starting point (z): The point from which the evidence accumulation process begins relative to the decision thresholds.
4. Non-decision time (T): The time required for stimulus encoding, motor preparation, and other non-decision-related processes.

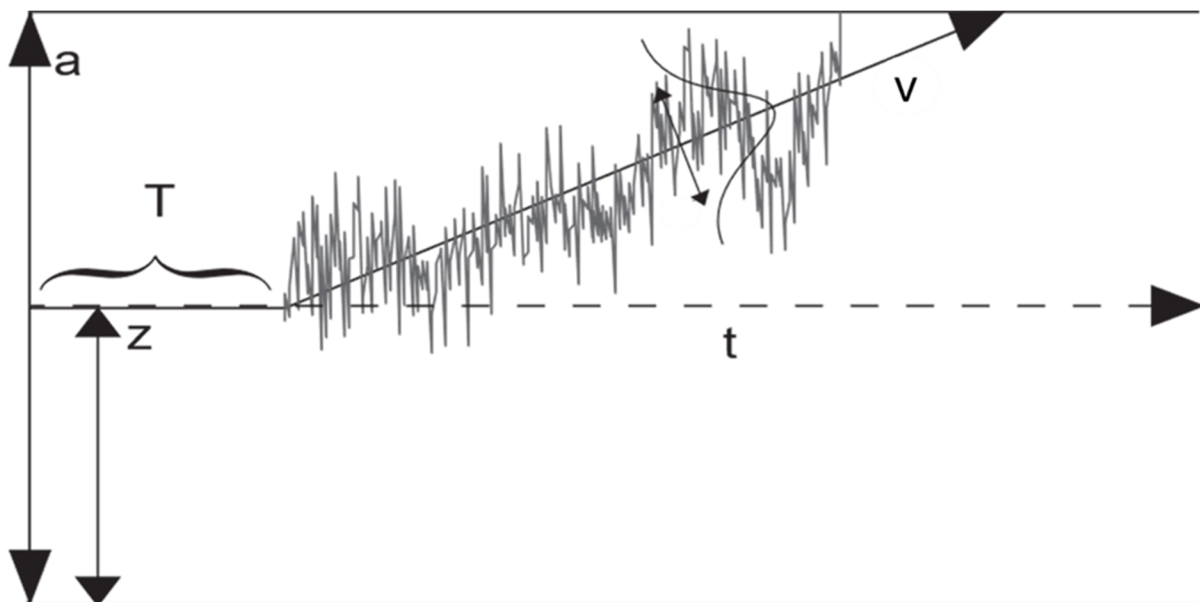


Figure 4.1. Depiction of a Drift Diffusion Model and its parameters

Previous research on perceptual decision making suggests that greater prospective confidence is associated with more cautious decision making (i.e., wider thresholds; Boldt et al., 2019).

However, low retrospective confidence and perceived errors in decisions have also been

associated with more cautiousness in subsequent decisions (Desender et al., 2019; Dutilh et al., 2012). Based on these mixed findings and our Study 1 finding of a parabolic relation between prospective confidence and response times, we expect a parabolic relation between prospective confidence and cautiousness. As our replication and extension of Study 1, we hypothesized that:

H2.4) Greater vs lower prospective confidence would be associated with greater retrospective confidence;

H2.5) Prospective confidence would motivate/demotivate deliberation on subsequent decisions—made apparent through response times and the threshold parameter of the drift diffusion model. Based on Study 1, we expected there to be an inverse-U relation between prospective confidence and response times (as well as the threshold parameter) such that participants are slowest and most cautious at a moderate level of confidence.

The hypotheses, method, data cleaning procedure, and analyses for Study 2 were preregistered on the Open Science Framework (<https://osf.io/g9ps4>).

Research Design

Procedure

Similar to Study 1, Study 2 consisted of a training phase, a measurement phase, and a test phase. Unlike Study 1, Study 2 was completed in person by participants from the PsychLing subject pool. Upon entry to the run room, participants were directed to sit at a computer with E-Prime software open and running the experiment. Following informed consent, the experimenter placed an Empatica E4 tracker on the non-dominant wrist of each participant and turned on the unit. Experienced electrodermal activity (EDA) was measured by electrodes placed adjacently on the wrist. Before the training phase, we assessed participants' metacognitive awareness (specifically,

participants' tendencies to monitor and evaluate their goal pursuit; Schraw & Dennison, 1994). Participants were then introduced to the task through three examples. In the training phase, participants were assigned to one of two conditions: *hindsight* (in which participants were shown questions with the correct answers indicated and had to indicate which they would have picked) or *foresight with feedback* (they answered questions and then were told the correct answer). There was no foresight without feedback condition because we were interested in affective responses to known outcomes. Participants completed 25 training trials, which consisted of 4 phases: a recovery period (5 seconds), a fixation period (2 seconds), the choice trial, and an affect reporting period (see Figure 4.2). The recovery period was to allow participants' EDA to return to a tonic baseline level following each trial (Markowitz et al., 2019). The baseline measures of EDA were assessed during the fixation before each trial to ensure that any increase or decrease in arousal represents a change relative to some standard. The analysis of the EDA data is outside the scope of the present dissertation. Following each trial, participants were asked to indicate their experienced affect toward knowing they were/would have been correct/incorrect.

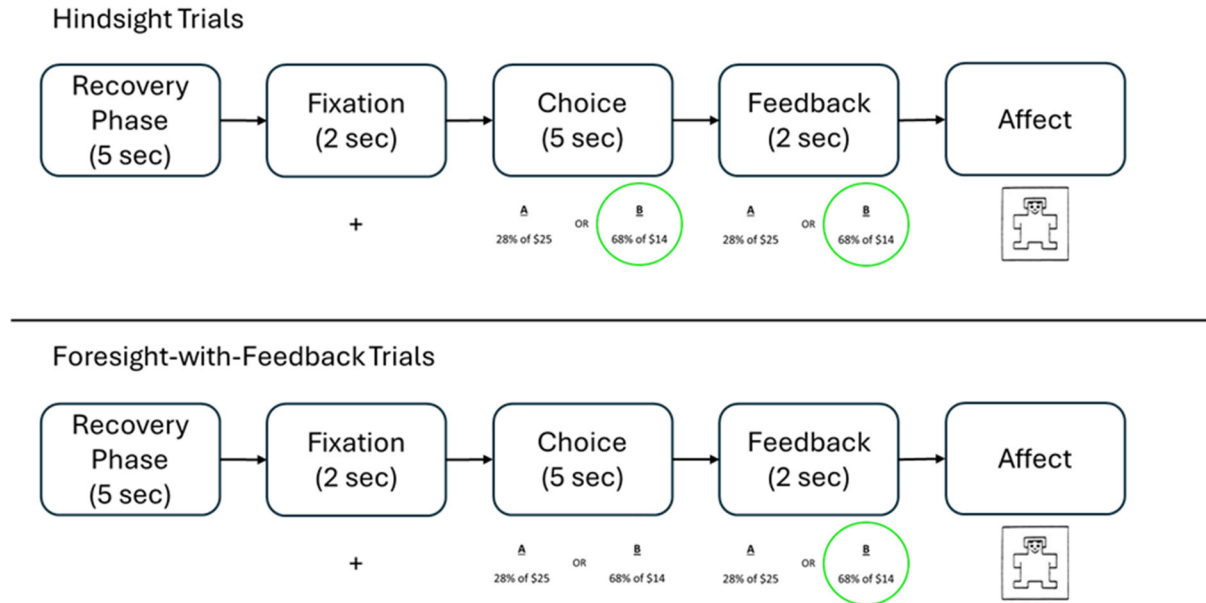


Figure 4.2. Depiction of each training trial. The overall timing of the training trials was kept consistent between the hindsight and foresight-with-feedback conditions.

In the foresight-with-feedback condition, participants had 5 seconds to select the larger of two mathematical expressions consisting of percentages and constants (e.g., 25% of 30 vs 10% of 70). Then, the correct answer was circled and remained on screen for 2 seconds. In the hindsight condition, participants had 5 seconds to indicate which expression they would have picked as larger. As in the foresight-with-feedback condition, the indicated answer remained on screen for an additional 2 seconds after the 5-second response period. In both conditions, there was a countdown timer at the bottom of the screen for each trial. After the training phase and before beginning the test phase, participants reported their prospective confidence for the test phase, task efficacy beliefs, and task difficulty beliefs as part of the measurement phase. We also measured numeric self-efficacy more broadly.

In the test phase, participants completed 45 trials of the same difficulty as the training trials. During this phase, all participants answered the trials without any feedback. After each choice, participants rated their retrospective confidence in their response. Following the test phase,

participants completed an objective numeracy measure. During the testing phase, we continued to collect experienced EDA using the Empatica E4.

Equipment

Participants were invited to complete the study in person in a room of desks that can seat up to eight participants, each with a desktop computer. The study was coded and prepared using E-Prime software, which afforded high-precision timing for matching up with EDA data and assessing response times. We had access to six Empatica E4 units, which allowed us to run up to six participants at a time.

Measures

Metacognitive Awareness. Participants answered a shortened version of the metacognitive awareness scale (Schraw & Dennison, 1994). The measure included 8 questions across the monitoring (e.g., “I ask myself if I have considered all options when solving a problem.”) and evaluation (e.g., “I ask myself if I have considered all options after I solve a problem.”) subscales. Participants responded on a 5-point Likert scale from Strongly Disagree to Strongly Agree. Items from each of the two subscales were averaged to create indices of monitoring and evaluation.

Attention Checks. After an introduction to the decision task, participants were asked two 3-alternative multiple-choice questions. The first is, "Missing responses will be..." The second is, "You will have ___ seconds to respond to each trial." The answers were “considered incorrect” and “5” respectively.

Baseline Prospective Confidence. Participants were asked to indicate their prospective confidence for the 25 training trials following an introduction to the task and before the training

trials. They were asked, “Now that you have some experience, we would like you to do some training for this task. We will now have you complete 25 training trials. Before you begin the training, how many of the 25 training trials do you think you will answer correctly?” Participants responded using a slider from 0 to 25. Participants were not aware of their assigned training condition. The value provided was transformed into a proportion by dividing the value by 25.

Training Choices (Manipulation Check). As a manipulation check, we calculated the proportion of correct responses for the foresight-with-feedback conditions as well as the proportion of responses participants indicated they would have gotten correct in the hindsight condition. Only trials participants responded to were included. Although the present study did not have a foresight condition, we would confirm that participants experienced hindsight bias if participants indicated the correct answer more often in the hindsight conditions than in the foresight-with-feedback condition (Hoch & Loewenstein, 1989). In Study 1, participants in the foresight and foresight-with-feedback conditions did not significantly differ in their accuracy for the training trials.

Affect to Outcomes. Following each training trial, we asked participants to indicate how they felt knowing they were correct/incorrect (in the foresight with feedback condition) or would have been correct/incorrect (in the hindsight condition). Affect was assessed using the valence and arousal scales of the self-assessment manikins (Bradley & Lang, 1994). For the valence scale, participants were told, “Select one character to describe how knowing whether you [were/would have been] right or wrong makes you feel from negative and displeased on the left to positive and pleased on the right.” The valence scale was scored from -2 (most negative manikin) to +2 (most positive manikin). For the arousal scale, participants were told, “Select one character to describe how knowing whether you [were/would have been] right or wrong makes you feel from

calm, drowsy, and peaceful on the left to excited, energized, and alert on the right.” The arousal scale was scored from 1 (calmest manikin) to 5 (most aroused manikin).

Participants’ average affective responses were also calculated across all training trials, trials for which participants indicated the correct response, and trials for which participants indicated the incorrect response. Participants’ affective sensitivity was calculated by subtracting participants’ average affect when indicating the incorrect response from their average affect when indicating the correct response.

Post-Training Prospective Confidence. Similar to Boldt and colleagues (2019), participants were asked to indicate their prospective confidence for the 45 test trials. They were asked, “Before you get started, how many of the following 45 trials do you think that you will answer correctly?” Participants responded using a slider from 0-45. The value provided was transformed into a proportion by dividing the value by 45.

Efficacy Beliefs. To assess efficacy beliefs for this task, participants were asked, “How bad or good do you think you are at this type of task?” after having responded to the training. Participants responded using a six-point semantic differential scale from “Extremely Bad” to “Extremely Good.”

Task Difficulty Beliefs. To assess difficulty beliefs for the trials, participants were asked, “How easy or difficult do you think this version of the task is for people in general?” Participants responded using a six-point semantic differential scale from “Extremely Easy” to “Extremely Difficult.”

Numeric Self-Efficacy. We assessed numeric self-efficacy using the ability subscale of the Subjective Numeracy Scale (Fagerlin et al., 2007; e.g., ‘How good are you at working with

fractions?'). Participants responded on 6-point Likert-type scales. Responses to the four items were averaged together.

Trial-By-Trial Retrospective Confidence. Following each trial in the testing phase, participants were asked to indicate their confidence in their immediately prior answer on a 0 (Certain Incorrect) to 100 (Certain Correct) scale as a measure of trial-by-trial retrospective confidence. Participants were asked, "How confident are you in your previous answer?" They were instructed how to use the slider: "Please use the slider below to indicate your belief of the probability that you are correct. For example, if you know that you didn't respond in time and are certain you are incorrect you would indicate 0 on the slider. If you completely guessed and have no clue, you would indicate 50 on the slider." The value provided was treated as a proportion of 100.

Overall Retrospective Confidence. Following the testing phase, participants were asked to indicate their overall retrospective confidence for the test trials. Participants were asked, "How many of the 45 trials do you think you answered correctly?" Participants indicated their responses using a slider from 0 to 45.

Objective Numeracy. Numeracy was assessed using the 4-item Numeric Understanding Measure (4-NUM), a measure that requires each participant to complete four math items (Silverstein et al., 2023).

Prospective Confidence Bias. Participants' accuracy in the 45 test trials was assessed as a proportion of correct responses from 0.0 to 1.0. The difference between reported prospective confidence and accuracy on the test trials was calculated. The resulting value varied between -1

and 1, with negative values representing underconfidence and positive values representing overconfidence relative to their actual performance.

Retrospective Confidence Brier Scores, Bias, and Sensitivity. Participants' calibration was calculated using Brier scores on the trial-by-trial retrospective confidence judgments (Brier, 1950). Brier scores quantify the average “distance” a participant is from perfect calibration (i.e., 100% confidence when correct and 0% confidence when incorrect). We also calculated participants' confidence sensitivity (i.e., the difference in trial-by-trial retrospective confidence when correct v. incorrect; often called slope) and confidence bias (difference between average confidence and true accuracy; often called accuracy; Yates, 1982).

Demographics. Participants were asked about their gender identity, age, and race/ethnicity.

Data Collection

Participants (N=230) were recruited from the PsychLing Subject Pool at the UO via Sona. We ran up to 6 participants per session. Participants received course credit for their participation.

Data Cleaning

To clean the data, first, we checked the attention checks. If participants answered both attention checks incorrectly, they were excluded. Next, any response greater than 3 standard deviations from the average response on a given question (aside from the trial-by-trial retrospective confidence questions) was considered missing for that question. Any participant missing more than 5% of the responses was excluded from any analyses. Finally, participants were excluded from any analyses if they failed to respond to 20% or more of the training or test trials. After cleaning, 199 participants remained (86.5% retained). One participant was removed for failing the attention checks, nine were removed for missing responses (other than the training trials), 7

were excluded for not responding to the test trials, and 14 were removed for failing to respond to the training trials.

Data Analysis

In general, participants were underconfident both prospectively and retrospectively. In the test phase, participants answered on average about 74% of the trials correctly. Participants' average prospective confidence was around 57%, and their average retrospective confidence was around 62%. Descriptive statistics can be found in Table 4.1, and bivariate correlations can be found in Table 4.2. Twenty-five participants (12.6%) indicated the correct answer for all the training trials. As such, their affect when incorrect and sensitivities could not be determined.

	Percent	Missing	Mean	Std. Deviation	Minimum	Maximum
Numeracy	-	0	1.65	0.87	0	4
Numeric Self-Efficacy	-	0	3.05	1.16	1	6
Monitoring	-	0	3.92	0.55	2	5
Evaluation	-	0	3.44	0.67	1	4.75
Baseline Prospective Confidence (Training)	-	0	0.53	0.17	.16	.88
Average Affect Valence	-	0	0.58	0.57	-1.60	2
Average Affect Valence to Positive Feedback	-	0	0.72	0.63	-1.60	2
Average Affect Valence to Negative Feedback	-	25	-0.25	0.72	-2	2
Valence Sensitivity	-	25	1.02	0.85	-0.57	3.14
Average Affect Arousal	-	0	2.72	0.74	1	4.72
Average Affect Arousal to Positive Feedback	-	0	2.72	0.79	1	4.74
Average Affect Arousal to Negative Feedback	-	25	2.83	0.80	1	5
Arousal Sensitivity	-	25	-0.08	0.73	-2.20	3.54
Post-Training Prospective Confidence (Test)	-	0	0.57	0.21	0.11	1
Prospective Confidence Bias (Test)	-	0	-0.17	0.18	-0.69	0.38
Task Self-Efficacy	-	0	3.61	1.15	1	6
Task Difficulty	-	0	3.28	0.97	1	5
Overall Test Accuracy	-	0	0.74	0.13	0.36	0.96
Average Response Time (Test)	-	0	3.24	0.49	1.77	4.02
Retrospective Confidence Bias (Test)	-	0	-0.11	0.15	-0.58	0.60
Average Retrospective Confidence when Correct (Test)	-	0	0.67	0.15	0.04	1
Average Retrospective Confidence when Incorrect (Test)	-	0	0.48	0.15	0.05	1
Retrospective Confidence Sensitivity (Test)	-	0	0.19	0.12	-0.06	0.48
Retrospective Confidence Brier Scores (Test)	-	0	0.19	0.08	0.06	0.60
Average Trial-by-Trial Confidence (Test)	-	0	0.63	0.15	0.04	1
Overall Retrospective Confidence (Test)	-	0	0.53	0.20	0.07	0.96
Age	-	2	19.34	1.60	18	30
Race/Ethnicity	White	66.3%	-	-	-	-
	Asian	6.0%	-	-	-	-
	Black/African American	2.0%	-	-	-	-
	Hispanic	8.0%	-	-	-	-
	Native Hawaiian/Pacific Islander	0.5%	-	-	-	-
	Other	1.0%	-	-	-	-
	Mixed Race	16.2%	-	-	-	-
	Gender					
Male	38.2%	-	-	-	-	-
Female	60.3%	-	-	-	-	-

Table 4.1. Descriptive statistics of Study 2 variables and demographics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
<i>Individual Difference</i>																												
1. Numeracy																												
2. Numeric Self-Efficacy	.40***																											
3. Monitoring	-.10	.01																										
4. Evaluation	-.12	.03	.56***																									
<i>Training Phase</i>																												
5. Baseline Prospective Confidence (Training)	.31***	.51***	-.02	-.06																								
6. Training Accuracy	.23**	.32***	-.07	-.05	.23***																							
7. Average Valence	-.11	.08	-.02	-.02	.06	.16*																						
8. Average Valence to Positive Feedback	-.13	.01	-.02	-.05	.02	-.13	.89***																					
9. Average Valence to Negative Feedback	-.18*	-.12	-.05	-.03	-.04	.02	.47***	.18*																				
10. Valence Sensitivity	.07	.16*	-.01	-.05	.08	-.05	.23**	.56***	-.71***																			
11. Average Arousal	-.07	-.01	.05	.11	-.17*	-.04	-.10	-.01	-.27***	.19*																		
12. Average Arousal to Positive Feedback	-.09	-.02	.05	.10	-.19**	-.09	-.12	.01	-.29***	.22**	.98***																	
13. Average Arousal to Negative Feedback	-.03	.05	.01	.06	-.11	.13	-.11	-.08	-.31***	.20**	.71***	.58***																
14. Arousal Sensitivity	-.05	-.06	-.01	-.00	-.04	-.19*	-.07	.05	.03	.01	.26***	.44***	-.48***															
15. Post-Training Prospective Confidence (Test)	.34***	.49***	-.12	-.15*	.71***	.39***	.25***	.19**	-.07	.23**	-.16*	-.16*	-.05	-.07														
16. Change in Prospective Confidence	.11	.10	-.14*	-.14*	-.16*	.28***	.28***	.24***	-.05	.22**	-.02	-.01	.06	-.04	.58***													
17. Prospective Confidence Bias (Test)	.15*	.28***	-.04	-.07	.51***	.16*	.27***	.25***	.09	.14	-.12	-.10	-.13	.07	.79***	.52***												
18. Task Self-Efficacy	.33***	.49***	-.19**	-.15*	.43***	.33***	.34***	.25***	.04	.18*	-.13	-.15*	.01	-.15*	.67***	.44***	.52***											
19. Task Difficulty	-.08	.06	.03	.07	-.08	-.07	.01	.02	-.03	.01	.17*	.17*	.06	.07	-.17*	-.15*	-.20**	-.15*										
<i>Testing Phase</i>																												
20. Overall Test Accuracy	.34***	.39***	-.13	-.14*	.41***	.40***	.01	-.06	-.23**	.16*	-.08	-.11	.10	-.20**	.47***	.18*	-.17*	.32***	.01									
21. Average Response Time	.14	.17*	-.02	-.11	.07	.08	-.04	.00	-.20**	.18*	-.01	-.00	-.06	.01	.07	.02	-.13	.02	.15*	.31***								
22. Average Trial-by-Trial Confidence	.34***	.54***	-.16*	-.15*	.60***	.33***	.20**	.11	.14	.05	-.20**	-.22**	-.14	-.02	.66***	.23***	.47***	.59***	-.11	.38***	.04							
23. Retrospective Confidence Bias	.04	.19**	-.05	-.03	.23**	-.01	.19**	.15*	.33***	-.10	-.13	-.12	-.22**	.16*	.24***	.07	.59***	.29***	-.11	-.47***	-.22**	.64***						
24. Average Trial-by-Trial Confidence when Correct	.35***	.55***	-.15*	-.13	.59***	.33***	.19**	.10	.11	.06	-.20**	-.22**	-.13	-.04	.64***	.23**	.48***	.57***	-.10	.35***	.04	.98***	.64***					
25. Average Trial-by-Trial Confidence when Incorrect	.09	.26***	-.13	-.12	.36***	.14*	.19**	.13	.28***	-.07	-.15*	-.13	-.18*	.12	.39***	.13	.41***	.38***	-.11	.04	-.03	.77***	.71***	.67***				
26. Retrospective Confidence Sensitivity	.32***	.37***	-.03	-.01	.29***	.23**	-.00	-.04	-.19*	.14	-.07	-.11	.04	-.19*	.32***	.12	.09	.25***	.01	.38***	.09	.26***	-.07	.42***	-.40***			
27. Retrospective Confidence Brier Score	-.48***	-.48***	.16*	.09	-.46***	-.36***	-.04	.03	.14	-.14	.10	.15*	-.07	.18*	-.56***	-.25***	-.21**	-.49***	-.00	-.60***	-.21**	-.64***	-.11	-.69***	-.21**	-.59***		
28. Overall Retrospective Confidence (Test)	.29***	.55***	-.18**	-.17*	.65***	.38***	.17*	.07	.05	.08	-.17*	-.17*	-.09	-.02	.79***	.35***	.59***	.62***	-.18*	.42***	-.00	.78***	.39***	.76***	.50***	.33***	-.54***	

Table 4.2. Study 2 Pearson correlations of between-subjects variables and indices. * indicates $< .05$, ** indicates $< .01$, *** indicates $> .001$.

Manipulation Check in the training phase. As a manipulation check, we compared participants in the hindsight condition and foresight-with-feedback condition on their average accuracy as a proportion (for participants in the hindsight condition, whether they reported they would have been accurate) using an independent samples t-test. As expected, there was a significant effect of condition ($t(197) = 3.80, p < .001, d = .54$), such that participants in the hindsight condition indicated they would have answered more trials correctly ($M = .83, SE = .02$) than participants actually answered in the foresight-with-feedback condition ($M = .75, SE = .01$). Thus, a modified hindsight effect emerged.

There are no significant effects of training condition ($t(197) = -1.36, p = .17, d = -.19$) for accuracy on the test trials.

Training Effects on Prospective Confidence, Efficacy, and Difficulty Judgments were largely similar. Following the training phase, participants in the foresight-with-feedback condition reported greater prospective confidence ($M = 0.60, SE = .02$) than participants in the hindsight condition, $M = 0.53, SE = .02, t(197) = 2.45, p = .02, d = .35$. Crucially, participants' baseline prospective confidence did not differ between conditions, $t(197) = -0.61, p = .54, d = -.09$. Participants in the foresight-with-feedback condition had a significant increase in prospective confidence from baseline following the training phase ($M = 0.08, SE = .02, t(107) = 5.56, p < .001, d = .54$), whereas participants in the hindsight condition did not significantly change from baseline following the training phase, $M = -0.01, SE = .01, t(90) = -0.37, p = .71, d = -.04$. Consistently, participants in the foresight-with-feedback condition reported greater efficacy beliefs following the training phase ($M = 3.92, SE = 0.11$) than participants in the hindsight condition, $M = 3.25, SE = 0.12, t(197) = 4.23, p < .001, d = .60$. However, participants' difficulty perceptions did not significantly differ by training condition, $t(197) = -1.09, p = .28, d$

= -.16. Thus, although hindsight participants thought they would have answered more correctly than foresight with feedback participants answered correctly, hindsight participants nonetheless reported lower prospective confidence and efficacy beliefs than foresight-with-feedback participants. This finding is inconsistent with what would be expected if participants were basing their prospective confidence solely on their perceived performance.

Training Conditions did not have significant effects on Retrospective Confidence. Regarding the testing phase, participants did not significantly differ by training condition in their average confidence in the testing phase ($t(197) = 1.10, p = .28, d = .16$), their average response times in the testing phase ($t(197) = 1.60, p = .25, d = .17$), their confidence accuracy (i.e., their over-/under-confidence; $t(197) = -0.09, p = .93, d = -.01$), and the sensitivity of their confidence to their accuracy, $t(197) = 0.29, p = .78, d = .04$. However, participants in the foresight-with-feedback condition were more calibrated in their confidence judgments as assessed through Brier Scores ($M = .18, SE = .01$) than in the hindsight condition, $M = .21, SE = .01, t(197) = -2.70, p = .007, d = -.38$. The lack of differences between conditions regarding retrospective confidence is consistent with the findings in Study 1; however, it is inconsistent with the observed differences in prospective confidence.

Hindsight weakened the association between confidence and accuracy between subjects.

Two separate linear regressions were fit predicting participants' post-training prospective confidence and participants' average trial-by-trial retrospective confidence from their accuracy on the test trials, the training condition, and the interaction of accuracy and condition. There was a significant interaction of training condition and accuracy predicting prospective confidence, $b = -0.60, SE = 0.20, t(195) = -3.02, p = .003; R^2 = .27, Adjusted R^2 = .26$. Participants who

received training in hindsight exhibited a weaker association between their prospective confidence and their accuracy ($b = 0.44$, $SE = 0.14$, $95\% CI: [0.16, 0.71]$) than participants who received training in foresight with feedback, $b = 1.04$, $SE = 0.14$, $95\% CI: [0.76, 1.32]$. See Figure 4.3. Moreover, there was a significant interaction of training condition and accuracy predicting average retrospective confidence, $b = -0.46$, $SE = 0.15$, $t(195) = -3.08$, $p = .002$; $R^2 = .19$, $Adjusted R^2 = .17$. Participants who received training in hindsight exhibited a weaker association between their prospective confidence and their accuracy ($b = 0.21$, $SE = 0.10$, $95\% CI: [.0002, 0.41]$) than participants who received training in foresight with feedback, $b = 0.67$, $SE = 0.11$, $95\% CI: [.46, 0.88]$. See Figure 4.4. Although participants' individual confidence sensitivities did not significantly differ between training conditions, hindsight appeared to reduce the association between confidence and accuracy between subjects. In other words, participants' confidence was less sensitive to their performance between subjects in hindsight compared to foresight with feedback. This pattern of results is somewhat consistent with the observation of Dawson and colleagues (1993) that physicians' confidence in their predictions was unrelated to their accuracy, which they attributed to repeated hindsight bias experiences. The question then becomes why did the foresight with feedback participants become more confident and calibrated relative to the hindsight participants?

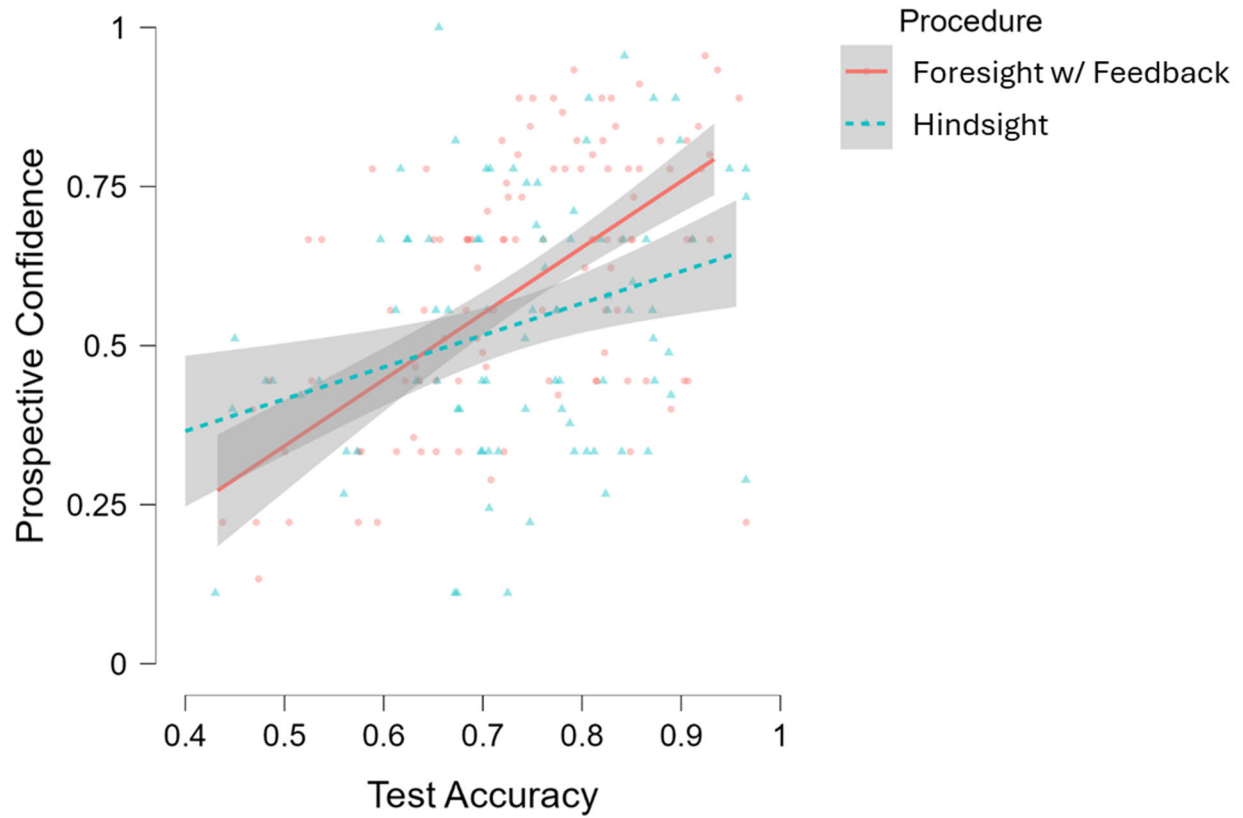


Figure 4.3. Expected marginal means of prospective confidence by test accuracy and training condition. Bands represent 95% confidence intervals.

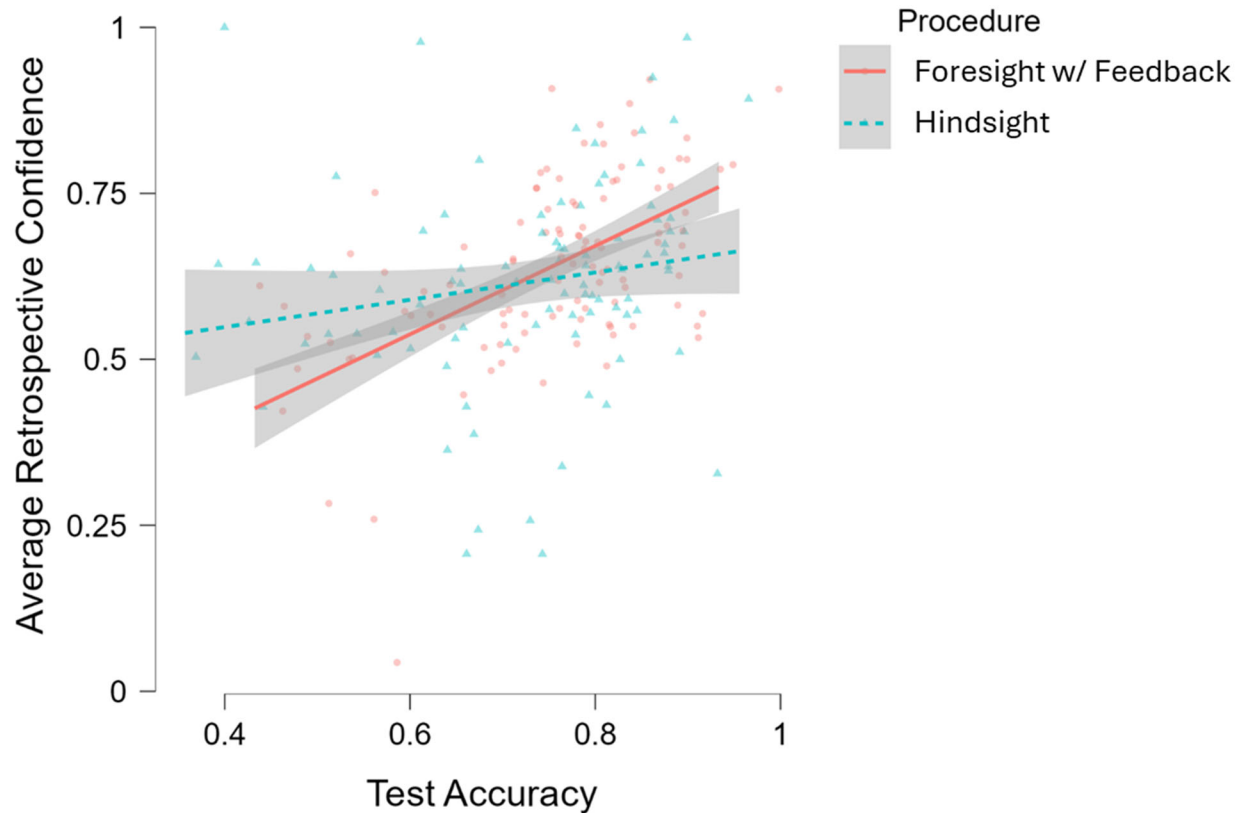


Figure 4.4. Expected marginal means of retrospective confidence by test accuracy and training condition. Bands represent 95% confidence intervals.

Hindsight muted Affective Responses to Feedback. To investigate participants' self-reported affective responses, separate hierarchical regressions were fit predicting participants' reported affect valence and arousal from their trial-by-trial accuracy and their training condition with random intercepts for each participant and each trial. We also explored whether participants' self-reported tendencies to monitor and evaluate their decision goal attainment moderated these effects.

Regarding participants' reported valence, there was a significant main effect of trial-level accuracy across both conditions such that participants felt more positive when they perceived they were/would have been correct ($M = 0.70$, $SE = 0.04$) than incorrect, ($M = -0.23$, $SE = 0.04$,

$F(1, 4695.2) = 1429.51, p < .001$), supporting H2.1 (that participants' affective responses would be sensitive to whether they were/would have been correct/incorrect). There was no significant main effect of training condition, $F(1, 212.1) = 0.04, p = .84$. However, there was a significant interaction between training condition and trial-level accuracy such that participants in the foresight-with-feedback were reported feeling more negative when incorrect but also more positive when correct than participants in the hindsight condition, $F(1, 4695.2) = 211.43, p < .001$. See Figure 4.5. Thus, consistent with H2.2a (that affective responses would be more sensitive to accuracy for foresight-with-feedback compared to hindsight participants) and the confidence/accuracy relationship above, participants' affective valence was more sensitive to their accuracy in the foresight-with-feedback condition than in the hindsight condition.

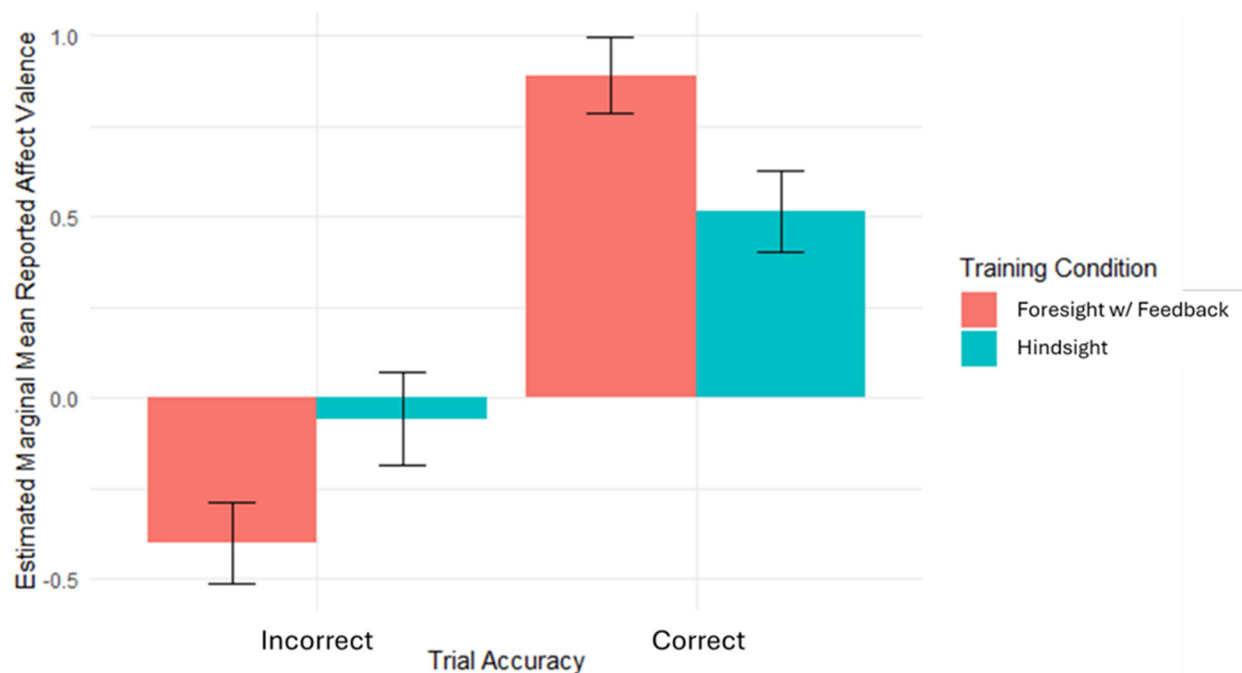


Figure 4.5. Estimated marginal means of affective valence by feedback and training condition. Error bars represent 95% confidence intervals.

Including metacognitive awareness, participants' self-reported tendency to evaluate their goal attainment was part of a significant three-way interaction with training condition and trial-level accuracy ($b = 0.41, SE = 0.08, t(4698.71) = -5.35, p < .001$). Specifically, greater evaluation tendency was associated with less negative affect valence when participants said they would have been incorrect in the hindsight condition ($b = 0.35, SE = 0.11, 95\% CI:[0.15, 0.56]$) but was not associated with valence when participants were incorrect in the foresight-with-feedback condition, $b = -0.03, SE = 0.08, 95\% CI:[-0.20, 0.13]$. See Figure 4.6. This finding is inconsistent with H2.2b (that greater metacognitive awareness would be associated with greater sensitivity of affect to accuracy); greater tendency to evaluate goal attainment appears to be associated with lower affect valence sensitivity for participants in the hindsight condition, specifically.

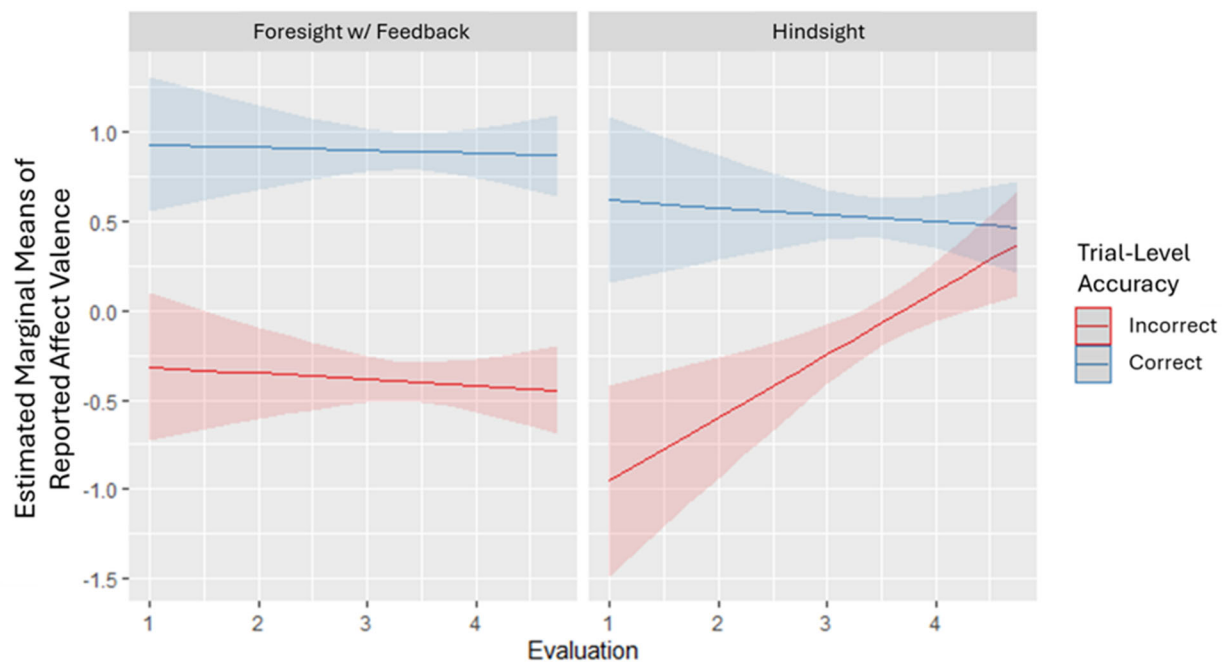


Figure 4.6. Estimated marginal means of affective valence by feedback, training condition, and evaluation tendency. Bands represent 95% confidence intervals.

Regarding participants' reported arousal, there was a significant main effect of training condition such that participants experienced greater affective arousal in the foresight-with-feedback

condition ($M = 2.85$, $SE = 0.07$) than in the hindsight condition, $M = 2.60$, $SE = 0.08$, $F(1, 207.4) = 5.90$, $p = .02$. There was no significant main effect of trial-level accuracy, $F(1, 4675.1) = 1.87$, $p = .17$. There was no significant interaction between training condition and trial-level accuracy, $F(1, 4675.1) = 0.49$, $p = .48$. Thus, participants' affective arousal was greater in the foresight-with-feedback condition than in the hindsight condition and was not sensitive to their accuracy. Participants' self-reported metacognitive awareness did not moderate any effects. See Figure 4.7.

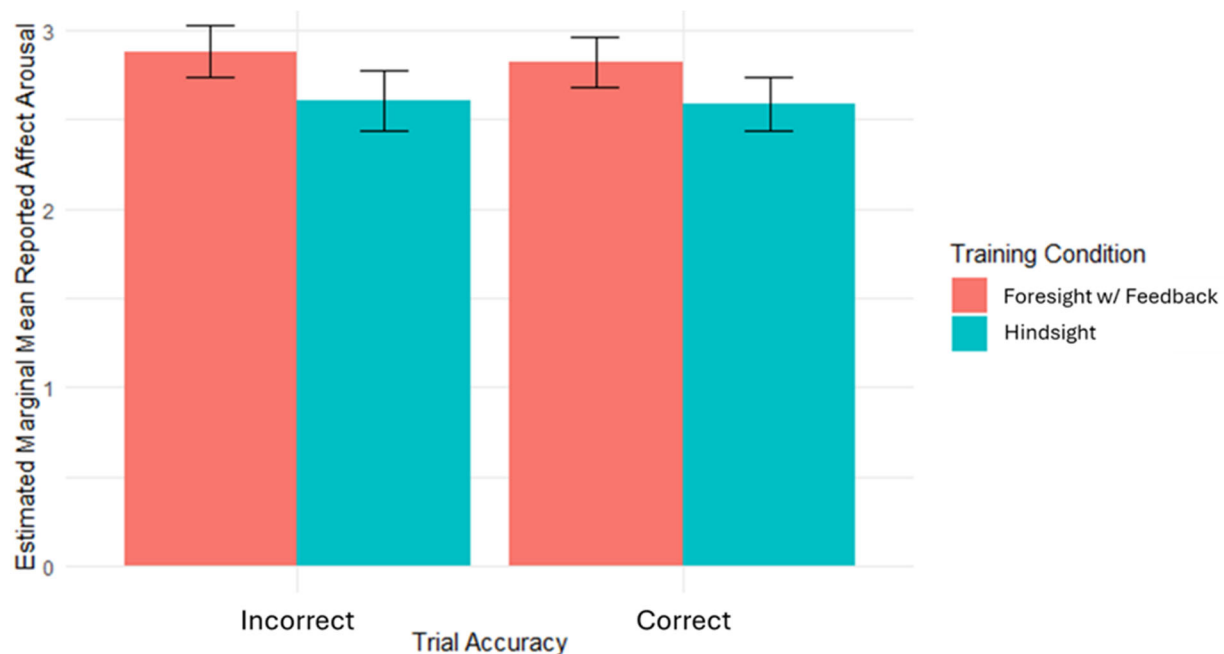


Figure 4.7. Estimated marginal means of affective arousal by feedback and training condition. Error bars represent 95% confidence intervals.

Affective Responses and Prospective Confidence. Next, we wanted to investigate whether affective responses informed prospective confidence. To test H2.3 (that valence and arousal of affective responses would predict changes in prospective confidence), we used participants' self-reported affective responses to predict their change in prospective confidence from baseline, controlling for their accuracy in the training phase. Specifically, we predicted participants' change in prospective confidence from their average valence and arousal to positive feedback,

their average valence and arousal to negative feedback, and their accuracy in the trials ($R^2 = .20$, *Adjusted R*² = .17). Participants who were more accurate in the foresight-with-feedback condition and participants who indicated they would have been more accurate in the hindsight condition during the training trials reported more positive change in prospective confidence after the training trials, $b = 0.40$, $SE = .08$, $t(168) = 5.09$, $p < .001$. Moreover, participants who reported feeling more positive affect to positive feedback also reported more positive change in prospective confidence after the training trials, $b = 0.07$, $SE = .02$, $t(168) = 4.04$, $p < .001$. See Figure 4.8. Participants' affective arousal and valence when incorrect did not significantly predict change in prospective confidence. Additionally, none of these predictors moderated each other; thus, no support existed for RQ2.1 (that participants with greater metacognitive awareness would change confidence more in line with affective responses). This finding is partially consistent with H2.3. Participants' affective valence, particularly to positive feedback, increased prospective confidence; however, affective arousal did not play a significant role. Moreover, affective responses to negative feedback had no significant effect on prospective confidence, suggesting an asymmetry in learning from positive and negative feedback.

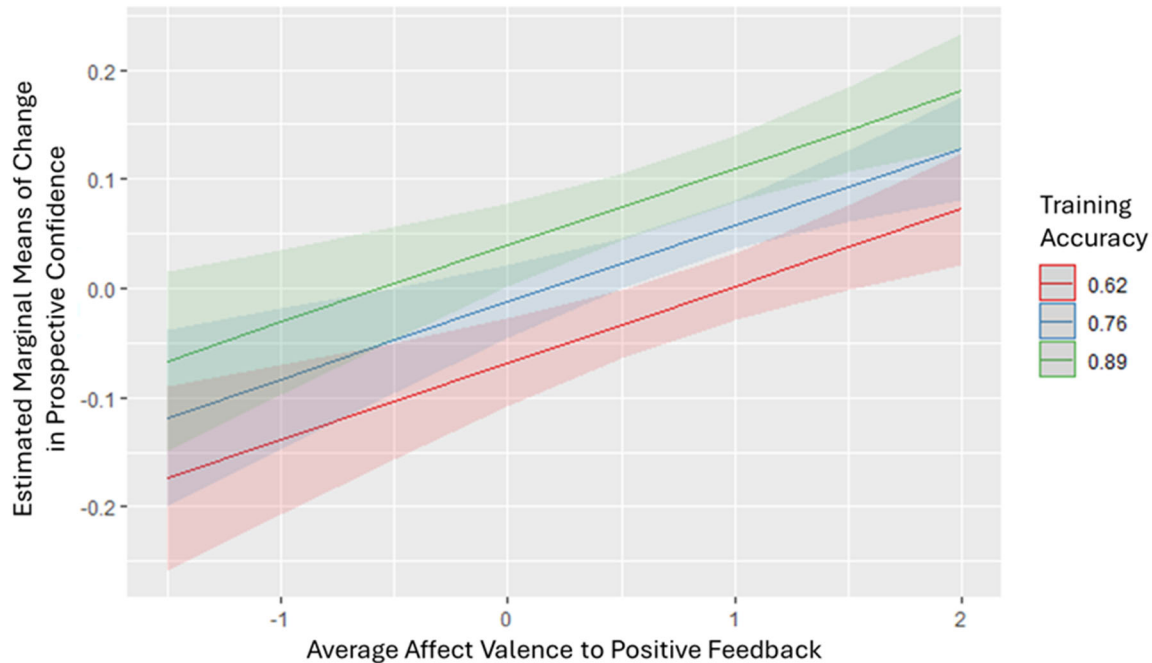


Figure 4.8. Estimated marginal means of change in prospective confidence by affect valence to positive feedback and training accuracy. Bands represent 95% confidence intervals.

Hindsight participants indicated that they would have been more accurate than foresight-with-feedback participants were actually accurate. However, hindsight participants also experienced less affective responses than foresight-with-feedback participants. Given that participants who perceived they were more accurate and participants who felt more positive toward positive feedback tended to increase in prospective confidence more, we investigated whether these processes conflict when participants gain experience in hindsight. We ran a mediation analysis to investigate how the training condition impacted participants' prospective confidence bias through training accuracy and participants' affective responses. We first specified a saturated parallel mediation model then non-significant pathways were removed to make a more parsimonious model ($\chi^2 (df = 1) = 2.30, p = .13, Robust CFI = .98, Robust TLI = .83, RMSEA = .08, SRMR = .02$). There were significant indirect effects of the training condition on prospective confidence bias through training accuracy ($ab1 = .05, z = 2.34, p = .02$) and affect valence to positive

feedback, $a_2b_2 = -.08$, $z = -3.01$, $p = .003$. The model results suggest that perceived accuracy and affective responses, particularly to positive feedback, both influence prospective confidence bias, but that hindsight had opposite effects on these factors. These cognitive and affective indirect effects effectively cancelled each other out concerning prospective confidence bias, $ab_{Total} = -.03$, $z = -1.91$, $p = .36$. See Figure 4.9.

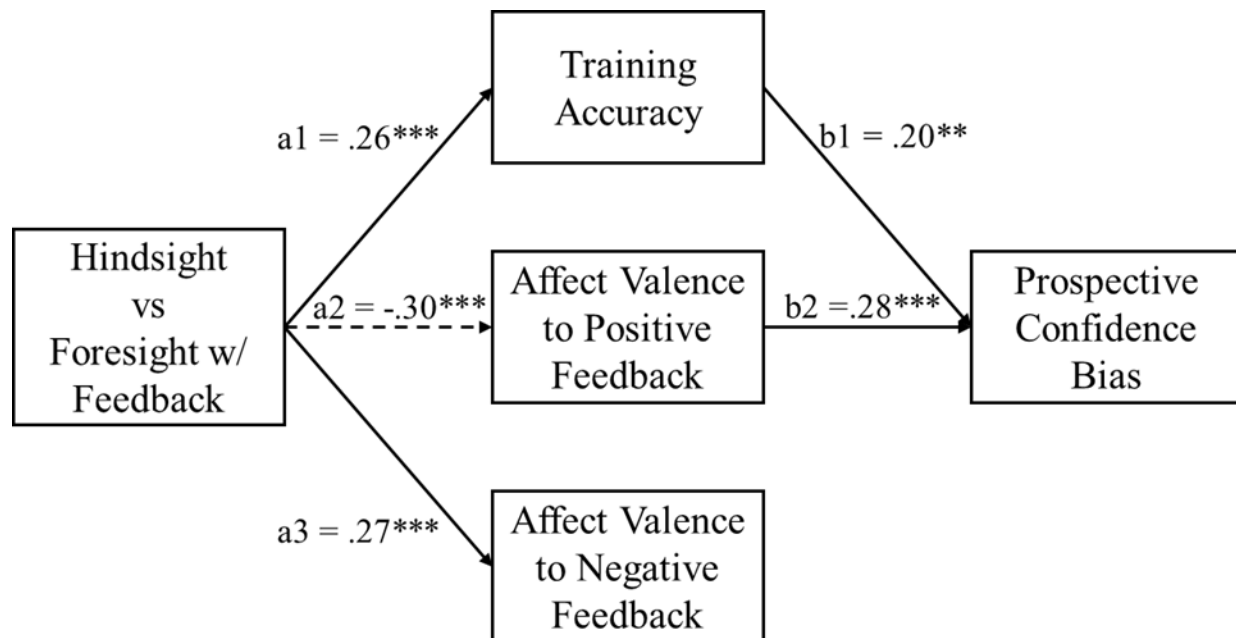


Figure 4.9. Parallel mediation model depicting indirect effects of training condition on prospective confidence bias (the difference between predicted and actual accuracy for the test phase) with standardized path coefficients. * indicates $< .05$, ** indicates $< .01$, *** indicates $> .001$.

Affective Responses and Retrospective Confidence. Next, we extended the mediation model to explore the effects on retrospective confidence and created a serial mediation model in which the training condition affected training accuracy and affect valence to positive and negative feedback, which affected prospective confidence bias, which in turn, affected retrospective confidence for correct and incorrect answers. We first specified a saturated serial mediation model and then removed non-significant pathways to make a more parsimonious model (χ^2 ($df =$

8) = 9.19, $p = .33$, *Robust CFI* = .99, *Robust TLI* = .99, *RMSEA* = .03, *SRMR* = .04; See Figure 4.10). See Table 4.3 for all significant indirect effects of training conditions on retrospective confidence. More positive affect to positive feedback informed retrospective confidence for correct and incorrect answers through prospective confidence. Moreover, more negative affect to negative feedback in training was associated with lower confidence judgments when incorrect in the test. These findings suggest that affective responses to negative feedback during training influence the extent to which participants' retrospective confidence judgments reflect their actual performance. Consistently, participants who experienced more positive affect when they were incorrect during training later showed lower sensitivity in their retrospective confidence judgments ($r = -.19, p = .01$). Overall, the indirect effects of the training condition on retrospective confidence for correct ($abc_{Correct} = .04, z = 1.56, p = .12$) and incorrect ($abc_{Incorrect} = .05, z = 1.71, p = .09$) answers cancelled out.

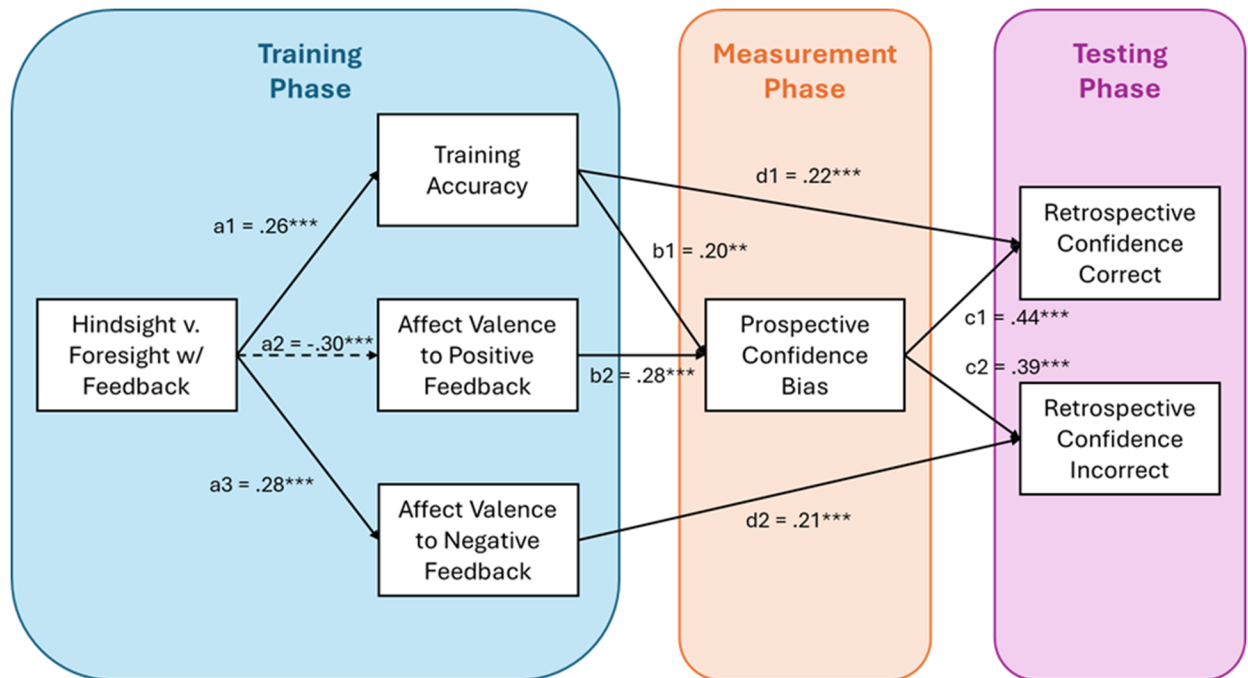


Figure 4.10. Serial mediation model depicting indirect effects of training condition on retrospective confidence for correct and incorrect answers with standardized path coefficients. * indicates $< .05$, ** indicates $< .01$, *** indicates $> .001$.

Path	Standardized		
	Estimate	Z-Value	p-Value
Training > Training Accuracy > Retrospective Confidence Correct	0.06	2.91	.004
Training > Affect Valence to Negative Feedback > Retrospective Confidence Incorrect	0.06	2.61	.009
Training > Training Accuracy > Prospective Confidence Bias > Retrospective Confidence Correct	0.02	2.23	.03
Training > Training Accuracy > Prospective Confidence Bias > Retrospective Confidence Incorrect	0.02	2.18	.006
Training > Affect Valence to Positive Feedback > Prospective Confidence Bias > Retrospective Confidence Correct	-0.04	-2.78	.005
Training > Affect Valence to Positive Feedback > Prospective Confidence Bias > Retrospective Confidence Incorrect	-0.03	-2.70	.009

Table 4.3. Significant indirect effects training condition on retrospective confidence for correct and incorrect answers.

Given the importance of affect and prospective confidence in retrospective confidence judgments, we sought to test if affective responses affected participants' retrospective confidence bias—the difference between average retrospective confidence and accuracy—and sensitivity. Thus, we modified the serial mediation model to explore the effects on retrospective confidence bias and sensitivity. In the resulting serial mediation model, the training condition affected training accuracy and affect valence to positive and negative feedback, which affected prospective confidence bias, which in turn, affected retrospective confidence bias and sensitivity. We first specified a saturated serial mediation model and then removed non-significant pathways to make a more parsimonious model ($\chi^2 (df = 7) = 6.12, p = .53, Robust CFI > .99, Robust TLI > .99, RMSEA < .001, SRMR = .03$; see Figure 4.11). Prospective confidence bias mediated the positive effects of affect valence to positive feedback and training accuracy on retrospective confidence bias. However, there was a direct effect of affect valence to negative feedback on retrospective confidence bias. Training accuracy and affect valence to negative feedback both directly influenced retrospective sensitivity. Overall, the indirect effects of the training condition on retrospective confidence bias ($abc_{Correct} = .04=3, z = 0.78, p = .44$) and sensitivity ($abc_{Incorrect} < .01, z < 0.01, p = .99$) cancelled out. See Table 4.4 for all significant indirect effects of training conditions on retrospective confidence bias and sensitivity.

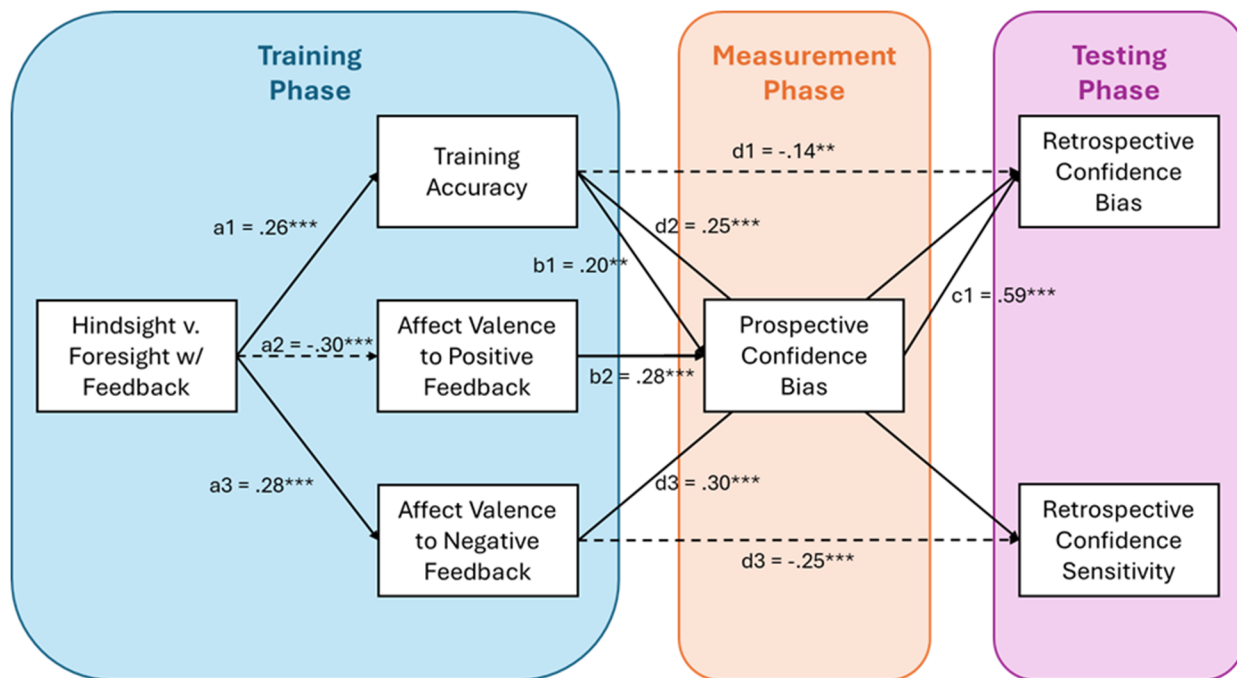


Figure 4.11. Serial mediation model depicting indirect effects of training condition on retrospective confidence bias and sensitivity with standardized path coefficients. * indicates < .05, ** indicates < .01, *** indicates < .001.

Path	Standardized Estimate	Z-Value	p-Value
Training > Training Accuracy > Retrospective Confidence Bias	-0.04	-2.12	.03
Training > Affect Valence Incorrect > Retrospective Confidence Bias	0.08	3.03	.002
Training > Training Accuracy > Retrospective Confidence Sensitivity	0.07	2.65	.008
Training > Affect Valence Incorrect > Retrospective Confidence Sensitivity	-0.07	-2.41	.02
Training > Training Accuracy > Prospective Confidence Bias > Retrospective Confidence Bias	0.03	2.28	.02
Training > Affect Valence Correct > Prospective Confidence Bias > Retrospective Confidence Bias	-0.05	-2.90	.004

Table 4.4. Significant indirect effects of training condition on retrospective confidence bias and sensitivity.

Based on Figure 4.11's results, experiencing negative affect toward perceived mistakes appears necessary for confidence to be sensitive to accuracy. Thus, we investigated whether affective valence toward negative feedback moderated the relation between accuracy and retrospective confidence. We fit a hierarchical linear model predicting trial-by-trial retrospective confidence from trial-level accuracy, participants' average reported valence to negative feedback, and the interaction of accuracy and valence to negative feedback (*Marginal R*² = .12, *Conditional R*² = .37). Participants were more confident when correct than when incorrect ($F(1, 7488.9) = 1029.62, p < .001$) and more confident when their average valence to negative feedback was more positive, $F(1, 223.7) = 17.49, p < .001$. Moreover, valence to negative feedback and accuracy interacted ($b = -0.04, SE = .01, t(7504) = -5.14, p < .001$), such that confidence was lower for incorrect responses when valence was more negative, ($b = 0.06, SE = 0.01, 95\% CI: [0.03, 0.09]$). See Figure 4.12. Additionally, we fit a linear regression predicting participants' average retrospective confidence from their overall accuracy on the test, participants' average reported valence to negative feedback in training, and the interaction of accuracy and valence to negative feedback ($R^2 = .22, Adjusted R^2 = .21$). There was a significant interaction between overall accuracy and valence to negative feedback ($b = -0.21, SE = .09, t(170) = -2.43, p = .02$), such that participants who reported less negative valence to negative feedback had a weaker association between their retrospective confidence and their overall accuracy. See Figure 4.13. Thus, participants who reported more negative affect to negative feedback exhibited greater retrospective confidence sensitivity both within and between subjects.

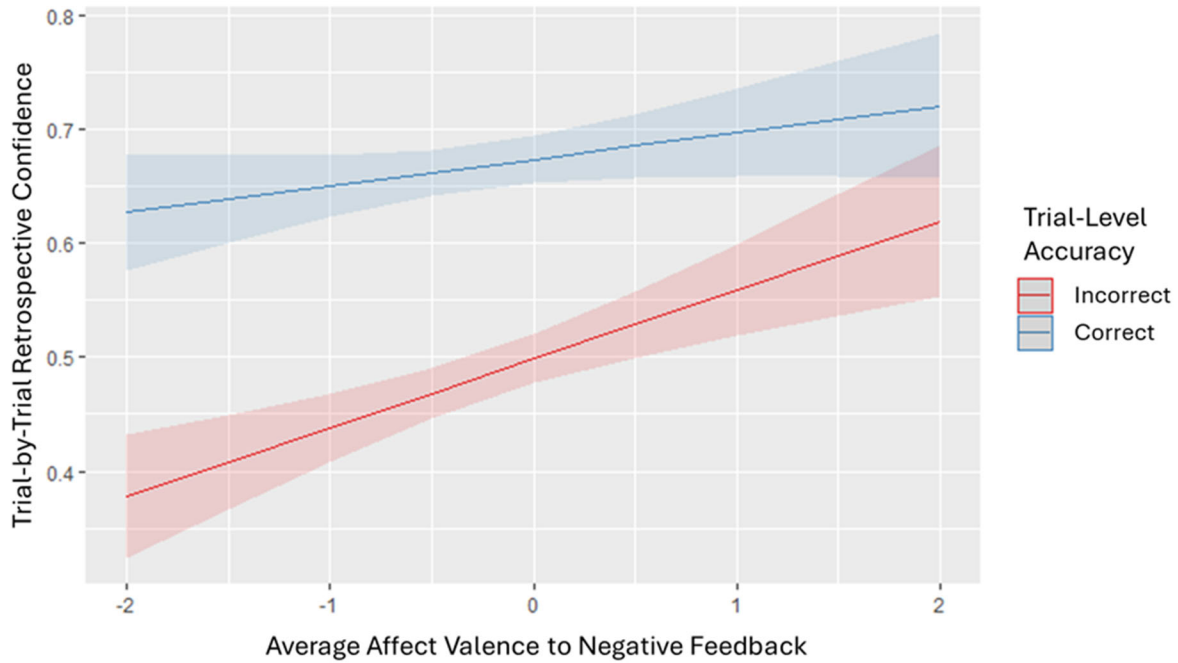


Figure 4.12. Estimates of marginal means of trial-by-trial retrospective confidence by trial accuracy and affective valence to negative feedback. Bands represent 95% confidence intervals.

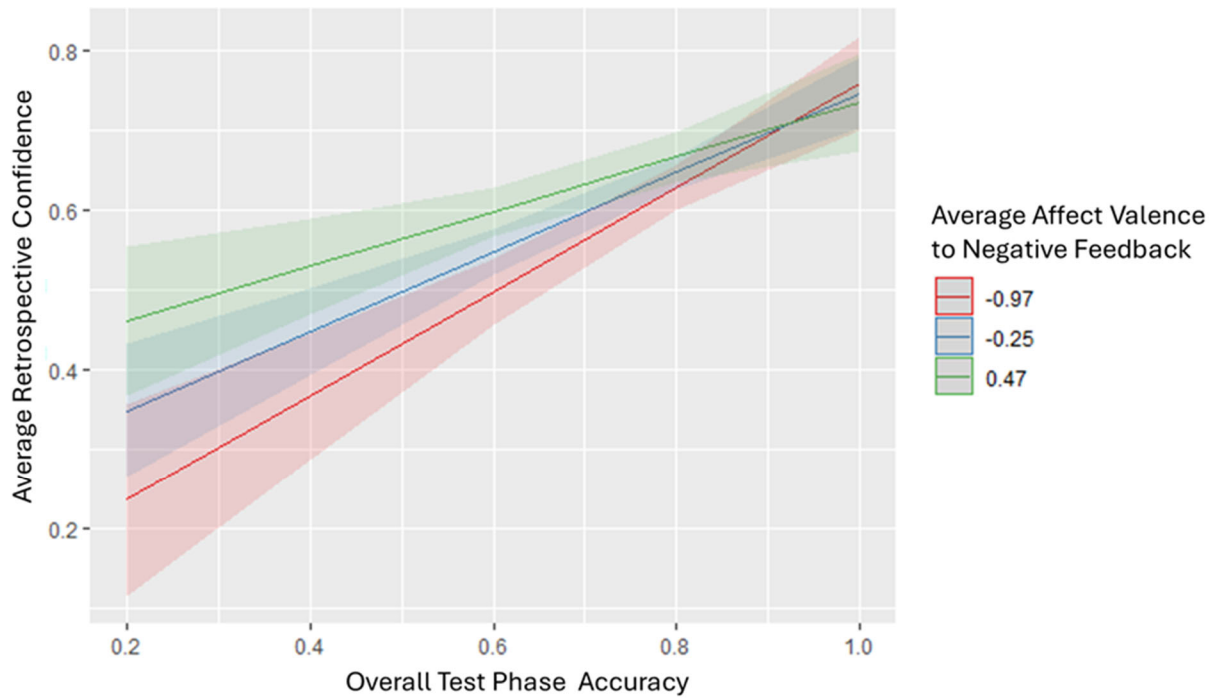


Figure 4.13. Estimates marginal means of average retrospective confidence by overall test accuracy and affective valence to negative feedback. Bands represent 95% confidence intervals.

Study 1 Models of Deliberation and Retrospective Confidence Largely Replicate. In Study 1, we found that prospective confidence predicted trial-by-trial retrospective confidence even when controlling for response times and accuracy (both at the trial level and participant level). This association supported the Confidence Cycle, which suggests that prospective confidence informs retrospective confidence. We sought to replicate the relation between prospective and retrospective confidence from Study 1 and test H2.4 (that greater prospective confidence would be associated with greater retrospective confidence). We explored whether we could predict trial-by-trial retrospective confidence judgments from the prospective confidence, trial-by-trial accuracy, trial-by-trial response times, the metacognitive awareness subscales, and overall accuracy as a proxy for ability, $Marginal R^2 = .25$, $Conditional R^2 = .47$. All independent variables other than accuracy were standardized. We found that greater prospective confidence predicted greater retrospective confidence ($F(1, 194.2) = 96.34, p < .001$; supporting H2.4), participants were more confident when they answered quickly than slowly ($F(1, 8657.1) = 117.45, p < .001$), and participants were more confident when correct than when incorrect ($F(1, 34390) = 8512.5, p < .001$). The metacognitive awareness measures did not significantly predict retrospective confidence. We also found a significant interaction between prospective confidence for the test and trial-by-trial response time ($b = -0.01, SE = .002, t = -3.64, p < .001$) such that prospective confidence was a stronger predictor of trial-by-trial retrospective confidence judgments for fast decisions than slow decisions. We also found a significant interaction between overall test accuracy and trial-by-trial accuracy ($b = 0.05, SE = .005, t = 11.13, p < .001$) such that participants who performed better on the test demonstrated greater sensitivity of their retrospective confidence judgments to their trial-by-trial accuracy.

In Study 1, we also found that prospective confidence had a parabolic relation with response time. This relation supported the Confidence Cycle, which suggests that prospective confidence modulates decision processes. We sought to replicate the relation between prospective confidence and response time from Study 1 and test H2.5 (that there is a parabolic relation between prospective confidence and response times). We tested whether we could predict trial-by-trial response times from prospective confidence (quadratic and linear terms), trial-by-trial accuracy, and overall accuracy (standardized) as a proxy for ability and the interaction of trial-level accuracy and overall accuracy, $Marginal R^2 = .04$, $Conditional R^2 = .23$. In addition to the quadratic and linear effects of prospective confidence ($b_{quadratic} = -1.75$, $SE = 0.68$, $t = -2.58$, $p = .01$; $b_{linear} = 1.84$, $SE = 0.78$, $t = 2.36$, $p = .02$), participants were slower at answering when correct ($M = 3.28$, $SE = .04$, $95\% CI: [3.19, 3.37]$) than when incorrect, $M = 3.03$, $SE = .05$, $95\% CI: [2.94, 3.12]$, $z = -11.26$, $p < .001$. There was also a significant interaction between trial-level accuracy and overall test accuracy ($b = 0.08$, $SE = 0.02$, $t = 3.69$, $p < .001$) such that participants who were overall more accurate were slower when they were correct than participants who were overall less accurate ($b = 1.27$, $SE = 0.28$, $95\% CI: [0.72, 1.82]$) but the relation between speed and overall accuracy was weaker when participants were incorrect, $b = 0.65$, $SE = 0.30$, $95\% CI: [0.07, 1.23]$, $z = 3.69$, $p < .001$. We observed an inverse-U-shaped relation between prospective confidence and response times consistent with H2.5 and Study 1.

Drift Diffusion Modeling. To investigate the effects of hindsight compared to foresight with feedback on decision processes, reaction times and choice data were analyzed using a Bayesian drift diffusion modeling (DDM) with regressors predicting random variation in drift rates (v ; the average rate of evidence accumulation supporting the correct v. incorrect answer) and threshold separation (a ; the amount of required to make a decision) using the HSSM package in Python

(Fengler et al., In Preparation). For the drift diffusion modeling, participants were excluded from analysis if they answered more than 90% of the test trials correctly to ensure the model would fit. Trials were excluded if they were faster than .30 seconds. This additional cleaning left us with 8,002 trials across 183 participants. Specifically, we predicted drift rates and threshold separation from the training condition (with hindsight coded as 1 and foresight-with-feedback coded as 0). Bayesian priors were specified for the regressors using a NUTS sampler, the z parameter was constrained to 0.50, and model convergence was assessed via R-hat and Monte-Carlo Standard Error (MCSE) diagnostics. The DDM indicated that drift rate (v) was significantly lower for participants in the hindsight condition compared to the foresight-with-feedback condition ($M_{Posterior} = -0.08$, 95% HDI: [-0.11, -0.04]). Moreover, the thresholds (a) were significantly narrower for participants in the hindsight condition compared to the foresight-with-feedback condition ($M_{Posterior} = -0.11$, 95% HDI: [-0.17, -0.05]). Convergence diagnostics showed R-hat \leq 1.01 and MCSE $<$.01 for all parameters. The hindsight training caused people to be less cautious and accumulate evidence more slowly than the foresight-with-feedback training. See Table 4.5 for all model parameter estimates.

<i>Parameter</i>	<i>Regressor</i>	<i>Mean (MCSE)</i>	<i>SD (MCSE)</i>	<i>95% HDI</i>	<i>R-hat</i>
<i>Starting Point Bias (z)</i>	-	0.50 (-)	-	-	-
	-	0.39 (0)	< 0.001 (0)	[0.39, 0.39]	1.00
<i>Non-Decision Time (T)</i>	Intercept	0.53 (0)	0.008 (0)	[0.52, 0.55]	1.00
	Hindsight (v. Foresight-with-Feedback)	-0.08 (.001)	0.02 (0)	[-0.11, -0.04]	1.00
<i>Drift Rate (v)</i>	Intercept	2.10 (.001)	0.02 (0)	[2.07, 2.14]	1.00
	Hindsight (v. Foresight-with-Feedback)	-0.11 (.001)	0.03 (.001)	[-0.17, -0.05]	1.01
<i>Threshold Separation (a)</i>	Intercept	0.53 (0)	0.008 (0)	[0.52, 0.55]	1.00
	Hindsight (v. Foresight-with-Feedback)	-0.08 (.001)	0.02 (0)	[-0.11, -0.04]	1.00

Table 4.5. Fitted DDM parameter estimates by training condition. The mean and standard deviation are for the posterior distribution. The starting point was constrained to 0.50.

Next, we wanted to investigate the relation between prospective confidence and the decision processes. To investigate the nature of the relation between prospective confidence, we allowed for a non-linear relation for prospective confidence given the observed parabolic relation between prospective confidence and response times. Specifically, we predicted drift rates and threshold separations from participants' prospective confidence, both squared and linear, to investigate the quadratic relation. The DDM indicated that drift rate (v) was exponentially higher for participants with greater prospective confidence ($M_{Posterior_Quadratic} = 0.91$, 95% HDI: [0.73, 1.00]; $M_{Posterior_Linear} = 0.03$, 95% HDI: [-0.15, 0.24]). Inconsistent with H2.5, there was an exponentially increasing relation between prospective confidence and thresholds' separation ($M_{Posterior_Quadratic} = 0.86$, 95% HDI: [0.60, 1.00]; $M_{Posterior_Linear} = 0.58$, 95% HDI: [0.32, 0.88]). Convergence diagnostics showed $R\text{-hat} \leq 1.01$ and $MCSE < .01$ for all parameters. Greater prospective confidence was associated with faster evidence accumulation (particularly for higher

levels of prospective confidence) and more cautious decision making. See Table 4.6 for all model parameter estimates. See Figure 4.14 for a depiction of the relations between prospective confidence and decision processes.

<i>Parameter</i>	<i>Regressor</i>	<i>Mean (MCSE)</i>	<i>SD (MCSE)</i>	<i>95% HDI</i>	<i>R-hat</i>
<i>Starting Point Bias (z)</i>	-	0.50 (-)	-	-	-
<i>Non-Decision Time (T)</i>	-	0.39 (0)	0.001 (0)	[0.39, 0.39]	1.00
<i>Drift Rate (v)</i>	Intercept	0.17 (.001)	0.04 (.001)	[0.10, 0.24]	1.00
	(Prospective Confidence) ²	0.91 (.003)	0.09 (.002)	[0.73, 1.00]	1.01
	Prospective Confidence	0.03 (.003)	0.10 (.002)	[-0.15, 0.24]	1.00
<i>Threshold Separation (a)</i>	Intercept	1.54 (.001)	0.05 (.001)	[1.45, 1.63]	1.00
	(Prospective Confidence) ²	0.86 (.006)	0.13 (.004)	[0.60, 1.00]	1.01
	Prospective Confidence	0.58 (.006)	0.14 (.005)	[0.32, 0.88]	1.01

Table 4.6. Fitted DDM parameter estimates by prospective confidence. The mean and standard deviation are for the posterior distribution. The starting point was constrained to 0.50.

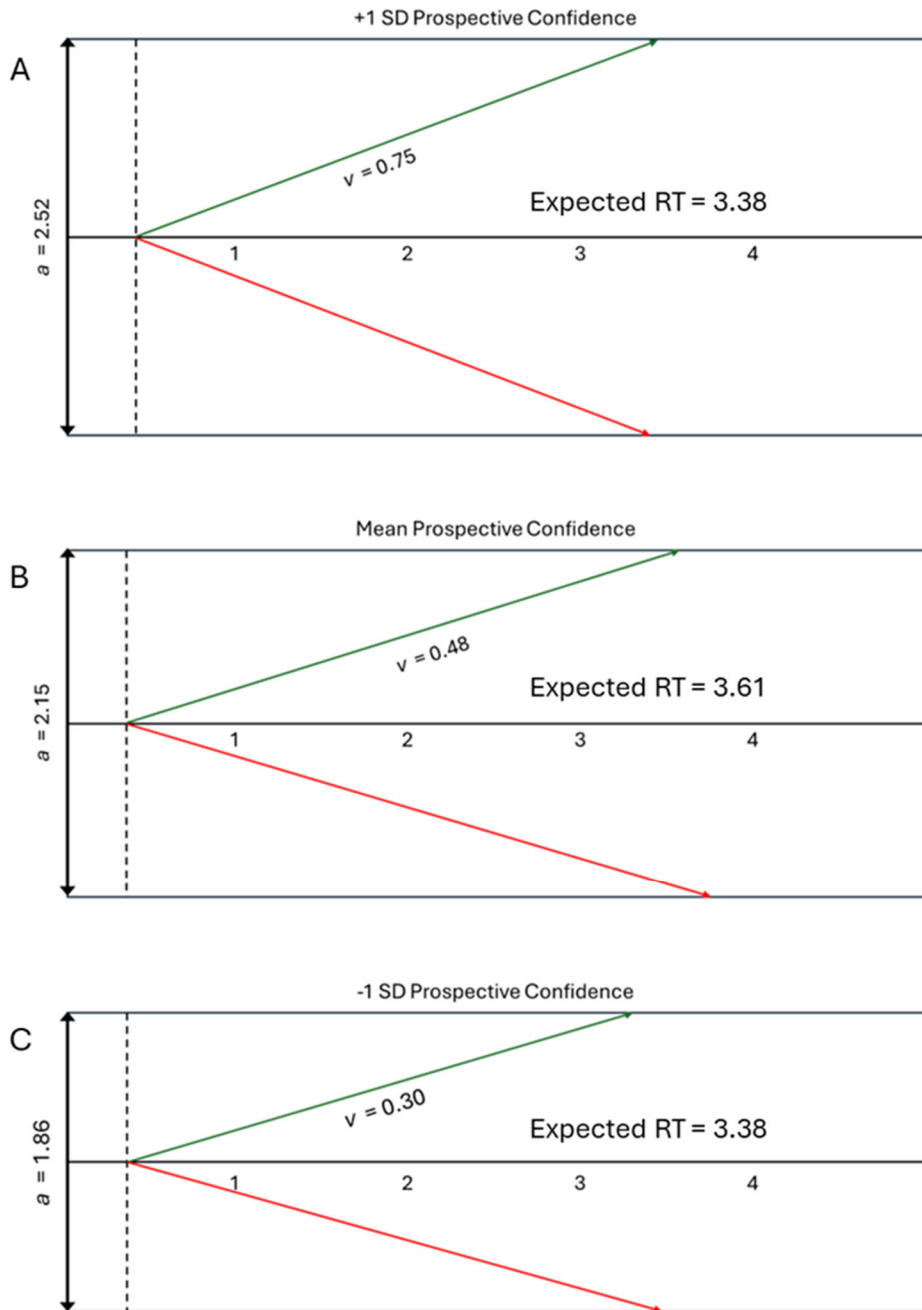


Figure 4.14. Depiction of estimated DDM parameters when prospective confidence is 1 standard deviation above average (A), average (B), and 1 standard deviation below average (C). The green arrow represents correct responses, and the red arrow represents incorrect responses. Expected response times (RTs) based on model simulations.

Discussion

As expected, participants in the hindsight condition indicated that they would have answered more items correctly than participants in the foresight with feedback conditions actually answered items correct, suggesting that we successfully induced hindsight bias (Hoch & Loewenstein, 1989). However, unexpectedly based on Study 1, participants in the hindsight training condition reported lower prospective confidence and task efficacy beliefs than participants in the foresight-with-feedback condition. In addition, participants in the hindsight condition did not show any significant change in their prospective confidence from baseline (as would have been expected from the hindsight literature), whereas their counterparts in the foresight-with-feedback condition increased in confidence. Further analyses suggested that the hindsight condition may have had little effect on prospective confidence because of competing effects—hindsight caused participants to perceive their accuracy for the training to be higher, but also resulted in relatively muted affective responses to feedback. Participants felt more positive to positive feedback than to negative feedback (H2.1 was supported), and the sensitivity of affect to the feedback was less for participants in the hindsight training condition (H2.2a was supported). Unexpectedly, participants who reported greater tendencies to evaluate their goal attainment indicated less sensitive affective valence responses in the hindsight condition only. Thus, H2.2b (that people with greater metacognitive awareness would have more sensitive affective responses) was not supported. More positive affect toward positive feedback and greater perceived accuracy were associated with increased prospective confidence. This finding partially supported H2.3 (that participants' affect valence and arousal would predict changes in prospective confidence). Affect toward negative feedback did not have a significant effect on prospective confidence, and arousal did not moderate the effect. Interestingly, although affective

responses to negative feedback did not appear to inform prospective confidence, they appear to directly influence retrospective confidence judgments when participants are incorrect.

Consistently, participants who reported more negative affect to negative feedback gave more sensitive retrospective confidence judgments. As expected (H2.4), greater prospective confidence and task-specific efficacy beliefs were associated with greater retrospective confidence, even when accounting for accuracy. We again observed a parabolic relation between prospective confidence and response times (H2.5). However, drift diffusion modeling revealed that hindsight participants required less evidence to decide and accumulated evidence more slowly.

Consistently, greater prospective confidence was associated with more cautious decision making and faster evidence accumulation.

The drift diffusion models shed light on how prospective confidence is associated with decision processes. In the Confidence Cycle, we assert that prospective confidence modulates decision processes. In Study 1, we showed a parabolic relation between prospective confidence and response times. The present study illuminates how prospective confidence could affect response times—both by increasing the rate of evidence accumulation and cautiousness. We demonstrated causal evidence that foresight-with-feedback (relative to hindsight) increases cautiousness and the rate of evidence accumulation in subsequent decisions. We suggest that these effects of the training conditions were due to the differences in prospective confidence causes. The results of our DDM are somewhat consistent with Boldt and colleagues (2019), who found that learned difficulty cues modulated decision threshold, with cues of easiness motivating more cautiousness. Moreover, they found that easy cues also caused greater preparatory neural activity, which could relate to our observed relation with the drift rate. However, our study is the first to

model the relations between prospective confidence and decision processes in a non-perceptual decision task. If we are correct that prospective confidence increased the rate of evidence accumulation and cautiousness in decisions, there could be an indirect effect of prospective confidence on retrospective confidence through decision processes. Indeed, faster drift rates and more cautiousness are both associated with greater retrospective confidence (see Pleskac & Busemeyer, 2010). However, it is possible that prospective confidence did not cause the observed changes in decision processes. It is also possible that the affective responses played a role in participants' decision processes, as well. We will discuss the potential role of affect in decision processes further in Chapter 5.

One explanation for the relatively muted affective responses for hindsight participants is the likely lack of surprise for participants in the hindsight condition. Mellers and colleagues (2013) posited that experiencing surprise amplifies affect. For example, McGraw and colleagues (2004) found that, when shooting basketballs, people felt more positive about making baskets and more negative about missing baskets when they were surprised by whether or not they made the basket. However, in hindsight, nothing is surprising because participants are not actually making the decisions and getting feedback. Mellers and colleagues (2013) predicted that because of this lack of surprise, both positive and negative events might become less emotional in hindsight. Our findings are consistent with their prediction. Moreover, our finding suggests that there are consequences for the lack of surprise, in that positive affect for success does not increase confidence, and negative affect for failures does not allow people to recognize when they are wrong.

As discussed in Chapter 1 (see *Decision Processes Inform Retrospective Confidence*), theories of

retrospective confidence can be categorized into information-based and experiential theories. Similarly, there appear to be cognitive (i.e., information-based) and affective (i.e., experiential) pathways to learning prospective confidence. Indeed, in the present studies, participants appeared to rely on their performance (whether real or judged in hindsight) to inform their prospective confidence but also relied on their affective responses, particularly to positive feedback. We find that training in hindsight had different effects on these two pathways. Regarding the cognitive pathway, hindsight increased participants' perceived performance, which was associated with increased prospective confidence (even when actual test performance was accounted for). This finding is consistent with classic assumptions that hindsight increases confidence by making people spuriously believe they are correct more (see Hawkins and Hastie, 1990; Dawson et al., 1993; Roese & Vohs, 2012). However, a different story emerged in the affective pathway. Participants' affective responses to feedback were relatively muted in the hindsight condition. Participants' more positive affect to positive feedback, in particular, appears to have led to greater prospective confidence for participants in the foresight-with-feedback condition compared to the hindsight condition. These competing pathways appear to have cancelled each other out in the present study when it came to the participants in the hindsight condition. The existence of two opposing pathways suggests that hindsight could increase or decrease prospective confidence relative to foresight, depending on the learning conditions. For example, if the task was more difficult in training, the effect of hindsight (versus foresight) on perceived accuracy might be greater (Winman et al., 1998). With a greater difference in perceived performance, the cognitive pathway could overcome the affective pathway to increase prospective confidence relative to foresight. Alternatively, if participants in foresight were made

to feel more positive about their answers and the task was easier, the affective pathway could overcome the cognitive pathway. In this case, foresight might increase prospective confidence more than hindsight.

When it comes to learning prospective confidence more broadly, it is interesting to consider the implications of the cognitive and affective pathways. The cognitive pathway is likely reliant on memory. As such, high cognitive load during the training might diminish working memory capacity and diminish the cognitive pathway relative to the affective pathway. Moreover, the cognitive pathway might be more susceptible to other memory biases, such as recency effects, and possibly memory distortions. This pathway might be less prone to bias when people are provided with veridical feedback. However, the role of the cognitive pathway is unclear when there is no feedback. Any cognitive pathway would likely be based entirely on retrospective confidence.

Conversely, the affective pathway might be susceptible to misattribution from various sources during training (Schwarz & Clore, 1983). For example, mood effects or emotions elicited by the task that are independent of performance could influence learned prospective confidence through the affective pathway (e.g., Wright & Mischel, 1982; Lyubomirsky et al., 2005). In the present study, we also observed an asymmetric influence of affect to positive and negative feedback on prospective confidence. Consistently, Salem-Garcia and colleagues (2023) suggest that confidence biases arise from learning biases in reinforcement learning contexts such as asymmetric updating. This reinforcement learning mechanism appears to be the case for the affective pathway for learning prospective confidence.

Even though affect to negative feedback did not influence prospective confidence (going into the

test phase), it did appear to inform retrospective confidence (during the test phase trials). In fact, participants who experienced more negative affect to negative feedback were less confident in their answers when incorrect, had more negative retrospective confidence bias, and demonstrated greater sensitivity in their retrospective confidence. These effects might also explain why participants in the hindsight condition were significantly less calibrated (in terms of Brier Scores) than those in the foresight-with-feedback condition. We suspect that affective responses are important for people to learn what it feels like to be wrong. These findings are consistent with the reasoning of Dawson and colleagues (1993) that hindsight bias can disassociate confidence from accuracy. These findings suggest that people need to feel wrong to learn and to be able to recognize what being wrong feels like. On the other hand, feeling good when right can increase prospective confidence.

We suggest that when participants receive negative feedback and experience negative affect in training, they begin to associate that negative affect with specific decision-making experiences that cue being wrong, such as slower responding and experienced disfluency, or with cues of task difficulty. This explanation is similar to Damasio's (1994) somatic marker hypothesis. When participants experience these cues in the testing phase, they might experience negative anticipatory affect and lower their confidence judgments accordingly. Indeed, people have been able to learn task difficulty cues and to change their prospective confidence based on those cues in perceptual decision tasks (Boldt et al., 2019). When those cues switched so that they no longer reflect difficulty, the cues still influenced prospective confidence and retrospective confidence judgments. Future research should investigate the role of affect in learning task difficulty cues and potential cues in the decision-making experience to inform retrospective confidence.

Moreover, future research could attempt to make affect to negative feedback more or less negative to investigate the effects on retrospective calibration.

Limitations

The present study was limited in several ways. First, the sample was relatively small (N=199). Our final sample was limited mostly by the expiration of the software needed to collect EDA data. This software expiration put a strict end date on our data collection. The small sample size limits our power for between-subjects analyses. However, the primary focus of the present study was on participants' affective responses during training, for which we had about 25 trials for each participant. By analyzing the effects of the training manipulation and feedback on the trial level, we were able to maintain a high level of power for investigating affective responses. However, power was weaker for between-subjects analyses, especially for interaction analyses. For example, we did not find any significant interactions between metacognitive awareness and affective responses in predicting prospective confidence (RQ2.1). It is possible that a larger sample would have provided enough power to find that participants with greater monitoring/evaluating tendencies relied on their affective responses more.

The present study did not include foresight without feedback as a condition. This exclusion was because the goal was to investigate the role of affect in learning from outcome information. When participants learn in foresight without feedback, they do not learn the correct answers. This lack of feedback does not mean that participants do not experience affective responses in foresight without feedback. We suspect that in this case, participants learn prospective confidence from their retrospective confidence and their affective responses to their perceived accuracy. Future research could assess the relationship between affective responses and

retrospective confidence in learning confidence. Without a foresight condition, we could not assess whether participants experienced hindsight as defined in Study 1—the difference in reported accuracy and actual accuracy between participants in the hindsight and foresight conditions, respectively. However, given that participants in the foresight and foresight-with-feedback conditions did not significantly differ in their accuracy in the training nor the testing phase of Study 1, we feel that the comparison between hindsight and foresight-with-feedback represents a reasonable proxy for assessing if participants experienced hindsight bias.

One potential explanation for why affective responses to positive feedback—but not negative feedback—were associated with prospective confidence is based on how we elicited prospective confidence judgments. Participants were asked the number of test trials they expected to answer correctly, which might have made participants overweight their affect when they were correct. It is possible that affect toward negative feedback could have predicted prospective confidence if it was elicited as the number of test trials participants expected to get incorrect. We cannot rule out this explanation. However, the finding that positive outcomes outweigh negative outcomes when learning confidence is not unique to our study. Salem-Garcia and colleagues (2023) similarly found asymmetric updating in reinforcement learning favoring positive outcomes and subsequently observed overconfidence for the same task. Moreover, Brown (1984) found that incidental positive affect increased attributions for successes to stable, internal factors (e.g., efficacy beliefs). Moreover, the present study cannot determine if asymmetric updating is a unique feature of the affective pathway or if it also occurs in the cognitive pathway. The present study did not vary the number of training trials and thus cannot dissociate between the amount of positive and the amount of negative feedback. Further research on reinforcement learning and

confidence might try to assess the cognitive and affective pathway separately to investigate if the asymmetry is unique to the affective pathway.

Finally, the present study is somewhat limited in its causal claims about hindsight and retrospective confidence. The manipulations did not yield many significant effects on retrospective confidence overall. However, mediation analyses suggested possible indirect effects from the manipulation through cognitive and affective pathways, prospective confidence, and ultimately to retrospective confidence. While most of the group comparisons did not yield significant differences, the groups did significantly differ in their Brier Scores, consistent with the mediation models.

Significance and Contributions to the Field

The present study highlights the role of affect in learning prospective confidence. In the proposed Confidence Cycle, we suggested that prospective confidence is learned from and updated by retrospective confidence. This study shed light on the process underlying how prospective confidence is learned, suggesting that there is a cognitive component as well as an affective component. Further, affective responses to negative feedback appeared to be particularly important for people to learn to recognize errors (thus, they are associated with greater sensitivity of retrospective confidence to accuracy), whereas affective responses to positive feedback were associated with greater prospective confidence. Moreover, there is a neurological basis for the role of affect in learning prospective confidence and learning to recognize errors. Bechara and colleagues (1997) found that healthy people (as opposed to patients with bilateral damage to their ventromedial prefrontal cortices [VMPFC]) developed affective responses to choosing “bad” decks in a gambling task (decks that would in the long run result in net losses) after only a few

negative outcomes. The healthy participants would experience anticipatory affect when choosing “bad” decks and start to avoid them, even without conscious awareness that the decks were bad. Patients with VMPFC damage, however, never developed these anticipatory affect responses, and, even if they could identify that the decks were bad, they would continue to pick the “bad” decks. Euston and colleagues (2012) suggest that the medial PFC learns and stores associations between events, context, locations, and corresponding emotional responses. Moreover, Lebreton and colleagues (2015) found that the VMPFC automatically integrates retrospective confidence in encoding values of choices, and multiple functional MRI studies have suggested that prospective confidence is associated with activity in the medial prefrontal cortex (Modirrousta & Fellows, 2008; Schnyer et al., 2004). Future neurological research should investigate if damage to the VMPFC impedes the learning of prospective confidence and retrospective confidence sensitivity by interfering with the integration of affect with outcome expectations.

The existence of an affective pathway for learning confidence may have important implications for improving calibration. When people are underconfident, it may be possible to address confidence bias by increasing positive affect toward being correct. When people are overconfident, reducing positive affect to correctness and increasing negative affect to incorrectness could reduce bias. If people’s confidence judgments are insensitive to their accuracy, increasing negative affect to incorrectness could help them learn how to identify errors in their decision making. Additionally, the asymmetric learning of prospective confidence from affect might help explain why people experience a “beginner’s bubble” (i.e., a rapid increase in confidence without a comparable increase in ability) when learning something new (Sanchez & Dunning, 2018).

Regarding hindsight bias, we find mixed support for the assumption that hindsight bias leads to overconfidence. The cognitive pathway appears to support the idea that hindsight increases confidence because people experience more spuriously positive feedback. However, the present study is the first known to support the speculation of Mellers and colleagues (2013) that affective responses are weaker in hindsight. Given that affect responses also appeared to inform prospective confidence, muting these responses lowers prospective confidence. In the present study, although participants who trained in hindsight appeared to have experienced hindsight bias (reporting greater accuracy than participants in the foresight-with-feedback conditions were actually accurate), they did not have any increase in prospective confidence from baseline. This lack of an increase challenges the long-standing assumptions that “Hindsight effects undoubtedly contribute to overconfidence...” (Hawkins & Hastie, 1990, p. 315).

V. GENERAL DISCUSSION

The pilot studies (Chapter 2) and two experiments (Chapters 3 and 4) in the present dissertation were designed to 1) investigate the relationship between hindsight bias and both prospective and retrospective confidence, 2) study how the amount of hindsight experience might affect changes in confidence, 3) begin to understand the consequences of prospective confidence and hindsight bias on confidence calibration and decision processes, and 4) investigate emotional processes underlying how people learn confidence from experiences. In Study 1, we found that participants who experienced training for a novel binary math choice task in hindsight (participants were given the answers and asked what they would have picked) reported greater prospective confidence than participants who experienced training in foresight without feedback (participants were never given the answers) but not greater than those who had received feedback. However, evaluating participants' changes from baseline, it appears that participants in the foresight condition decreased in their prospective confidence, while participants in the hindsight and foresight with feedback conditions remained steady. Thus, despite confirming that participants in the hindsight condition experienced a hindsight bias, we did not find any *increase* in prospective confidence due to the hindsight condition. Thus, GH1 (that repeated hindsight experiences vs experiences in foresight [without feedback] would increase prospective confidence) was only partly supported. The conditions had no significant effect on participants' more general numeric self-efficacy despite having effects on the task-specific efficacy beliefs. There were mixed results of the effect of the training length on prospective confidence. The hypothesized interaction of training length and condition (GH3: that more vs fewer hindsight experiences would increase prospective confidence) emerged only when assessing participants' change in their prospective confidence from baseline; however, the pattern of results was more consistent with a greater

number of trials in the foresight condition lowering prospective confidence rather than more trials in the hindsight condition raising prospective confidence. Thus, GH3 was not supported. Training in hindsight (v. foresight) increased retrospective confidence in subsequent decisions (supporting GH2: that repeated hindsight experiences vs experiences in foresight [without feedback] would increase retrospective confidence). However, this effect was diminished for trial-by-trial judgments when participants decided slowly. There were no significant effects of training length on retrospective confidence and thus no support for GH4 (that more vs fewer hindsight experiences would increase retrospective confidence). Finally, we observed a parabolic relation between prospective confidence and response times, suggesting that prospective confidence might have a non-linear relation with deliberation (supporting GH5: that increases in prospective confidence will motivate/demotivate deliberation on subsequent decisions).

Participants with greater tendencies to monitor their goal attainment were less underconfident prospectively in the hindsight condition, generally, and in the foresight condition when training was short. However, when training was long, participants in the foresight condition who reported greater monitoring tendencies were more underconfident (GRQ2: whether metacognitive awareness would moderate effects of training conditions on prospective and retrospective confidence). Retrospectively, participants who reported greater monitoring were less underconfident following training in the foresight condition when it was short, while participants who reported greater monitoring were less underconfident following training in the hindsight and foresight-with-feedback conditions when it was long. Put differently, high monitors were generally more accurate in their confidence judgments following short training in foresight and long training in hindsight and foresight with feedback. Participants did not significantly differ by

condition in their confidence sensitivity (GRQ1: whether hindsight would reduce calibration and sensitivity of retrospective confidence).

In Study 2, we generally replicated the findings that prospective confidence predicts retrospective confidence (supporting GH2) and the parabolic relation between prospective confidence and response times (supporting GH5). Moreover, drift diffusion modeling revealed that changes in response times were the consequences of changes in both the rate of evidence accumulation and the amount of evidence required to decide. Somewhat surprisingly, participants in the foresight-with-feedback condition reported greater prospective confidence than those in the hindsight condition. Similar to Study 1, participants in the hindsight condition did not significantly change in prospective confidence due to the training. However, prospective confidence increased during training for participants in the foresight-with-feedback condition. Participants' affective responses were generally more muted in the hindsight condition than the foresight-with-feedback condition. These affective responses appear critical to how participants learn prospective confidence (supporting GH6: that affective responses to feedback during training would predict changes in prospective confidence), and for participants to be sensitive to errors. Participants in the hindsight condition exhibited a weaker relation between their average accuracy and confidence, both for their prospective and retrospective confidence judgments, than participants in the foresight-with-feedback condition. Moreover, participants in the hindsight conditions were significantly less calibrated in their retrospective confidence judgments assessed by Brier scores (GRQ1).

The results of the present dissertation offer important insights into the effects of hindsight bias, prospective confidence, and affective responses for subsequent confidence judgments and

decision processes. First, although participants in the hindsight condition reported greater prospective confidence than those in the foresight (without feedback) condition, these effects appeared largely due to decreases in the foresight condition rather than increases from hindsight bias itself. As such, the present dissertation challenges the speculation that hindsight bias “undoubtedly” causes overconfidence (Hawkins & Hastie, 1990; see *Consequences of Hindsight for Subsequent Confidence Judgments* below). Second, prospective confidence appeared to play a motivational role for decision processes and an informational role for retrospective confidence judgments, consistent with the proposed confidence cycle (see *Functions of Prospective Confidence* below). Finally, emotional responses during training appeared to play a critical role in the learning of confidence: affective responses to positive feedback predicted shifts in prospective confidence, and affective responses to negative feedback predicted how sensitive participants' retrospective confidence judgments were to whether their decisions were correct or incorrect (see *Affect and Confidence* below).

Consequences of Hindsight for Subsequent Confidence Judgments

Dawson and colleagues (1993) found that, when it came to estimating patients' hemodynamic status without performing right heart catheterization, more experienced participants were more confident but no more accurate than less experienced physicians. Additionally, they observed no relation between confidence and accuracy. Dawson and colleagues theorized that physicians repeatedly experienced hindsight bias when assessing hemodynamic status and that more experiences of hindsight bias increased confidence without an opportunity to improve accuracy. Our findings were partially consistent with this explanation of the findings of Dawson and colleagues. Experiencing hindsight bias during training did appear to have consequences for

subsequent confidence judgments. Relative to participants in the foresight condition, participants in the hindsight condition came out of the training trial with greater prospective confidence for the subsequent test trials. Moreover, participants in the hindsight condition also had greater retrospective confidence during the test trials, particularly for fast decisions. In Study 2, we found that relative to participants in the foresight-with-feedback condition, participants in the hindsight condition had a weaker relation between their confidence judgments and their accuracy, interpersonally. This finding was somewhat consistent with Dawson and colleagues' (1993) observed lack of a relation between confidence and accuracy. Inconsistent with the findings and theorizing of Dawson and colleagues, we found no effect of the amount of experience on participants' prospective and retrospective confidence. Additionally, participants in the hindsight condition *did not show an increase* in prospective confidence. Rather, participants in the foresight condition *decreased* their prospective confidence following the training. To our knowledge, Study 1 is the first study to investigate the consequences of experiencing hindsight bias for subsequent confidence judgments for similar decisions.

Does hindsight bias cause overconfidence? The pattern of results in Study 1 was more consistent with the hindsight condition protecting participants from losing confidence. However, participants were, on average, underconfident for the task. Had participants' confidence matched their accuracy initially, we suspect that foresight participants would not have decreased in prospective confidence, and the hindsight participants may have increased in prospective confidence. Future studies should develop a task for which participants were not underconfident to investigate whether hindsight bias increases confidence for subsequent decisions in that case. Additionally, the protective effect of the hindsight condition (compared to the foresight

condition) was short-lived. Following the training, participants in the hindsight and foresight-with-feedback conditions had comparable levels of prospective confidence. However, after one week, participants in the hindsight condition did not significantly differ in their prospective confidence from those in the foresight condition; participants in the foresight-with-feedback condition maintained their levels of prospective confidence. Overall, we did not show significant effects of the number of training trials. It is possible that participants considered the experience as a whole or attended to some subset of trials (e.g., they may have considered only the peak and end; Kahneman et al., 1993). Future research should investigate whether having multiple distinct sessions would cause there to be effects of the amount of experience rather than varying the length of a single session. As they stand, our present findings call into question the long-held assumption that “Hindsight effects undoubtedly contribute to overconfidence...” (Hawkins & Hastie, 1990, p. 315). Although we suspect that hindsight could cause overconfidence, our findings suggest that it is not a given and that the effects of hindsight on confidence might be short-lived.

Can hindsight bias be leveraged to study the effects of prospective confidence? We found that participants in the hindsight condition rated themselves as better at the task than those in the foresight condition. This finding was consistent with the findings of Bradfield and Wells (2005) that positive feedback caused participants to distort the quality of their decision making and see themselves as better at making similar decisions. However, we did not observe any significant effects on more general numeric self-efficacy. The present studies employed a numeric decision task that was very simple; this methodological choice was more conducive for evaluating response times, but may have limited our ability to alter the more general numeric self-efficacy.

Specifically, all the trials used the same type of questions and were very simple so they could be answered relatively quickly. To manipulate numeric self-efficacy more broadly, a wider variety of items might be useful. If more varied training items are used, affording participants hindsight for difficult math problems might be an effective way to manipulate numeric confidence relative to having participants complete the problems without feedback. Using hindsight as a manipulation might allow all participants to see the same math problems while still manipulating confidence.

Functions of Prospective Confidence

While other studies have investigated the impact of prospective confidence on retrospective confidence and decision making (e.g., Boldt et al., 2019; Ehrlinger & Dunning, 2003), this dissertation represents the first study to do so at the level of individual decision and using non-perceptual decisions. Across Studies 1 and 2, we demonstrated that biasing prospective confidence through training biased subsequent retrospective confidence. Interestingly, prospective confidence only predicted trial-by-trial retrospective confidence after controlling for the effect of response times, suggesting that their judgments may rely on multiple cues (consistent with Koriat, 2007). However, overall retrospective confidence (judged at the end of the testing phase) was more closely related to prospective confidence and the training condition. The relation between prospective confidence and overall retrospective confidence was consistent with the finding of Ehrlinger and Dunning (2003)—that efficacy beliefs informed judgments of overall performance on a test. One possible explanation for why the relation between overall retrospective confidence and prospective confidence is closer than that between trial-by-trial retrospective confidence and prospective confidence is that the trial-level judgment takes the

specific decision-making experience more into account. On the other hand, participants' more holistic retrospective confidence judgment might be based more on their prospective confidence. However, it is also possible that the more similar response scales are responsible for the closer relationship (prospective and overall retrospective confidence were indicated using sliders from 0-45, while trial-by-trial retrospective confidence was indicated using sliders from 0-100). However, challenging the response scale explanation, a similar pattern was observed for the relations between efficacy beliefs and the two types of retrospective confidence judgments—efficacy beliefs are more strongly associated with overall retrospective confidence than trial-by-trial retrospective confidence, despite efficacy being assessed using a six-point semantic differential scale. Future work could try to distinguish between how retrospective confidence for individual decisions versus a series of decisions is judged and the roles of each in how people learn prospective confidence from retrospective confidence. In Study 1, participants' reported prospective confidence and efficacy beliefs after a week were more closely associated with their overall retrospective confidence than their average trial-by-trial retrospective confidence, suggesting that the more holistic judgments might matter more for updating prospective confidence.

We find that, in addition to impacting subsequent retrospective confidence judgments, prospective confidence might influence *how* people decide. Across Studies 1 and 2, we found that prospective confidence had a parabolic relation with response times—participants tended to make decisions slower when they had moderate levels of prospective confidence. In Study 2, we employed drift diffusion modeling to investigate why response times varied by prospective confidence. Greater prospective confidence was associated with greater cautiousness (wider

evidence thresholds) and faster rates of evidence accumulation (drift rates). The finding that greater cautiousness was associated with greater prospective confidence was consistent with some prior findings (e.g., Boldt et al., 2019; who found more cautious decision making for perceptual decisions cued to be easy) but somewhat inconsistent with others (e.g., Zhao et al., 2022; who found more cautious decision making when disfluent text [believed to cue difficulty] was used). Future work should continue to investigate what motivates more cautious decision making and the role of prospective confidence. It is possible that other factors, such as mindset (growth v. achievement) and incentives, play a role in whether prospective confidence motivates or demotivates cautious decisions. It is also possible that there are different effects of trait-level and state-level prospective confidence (Hamaker, 2023). It is possible that higher trait-level prospective confidence increases cautiousness and state-level decreases in confidence also motivate more cautious decisions making. These effects of trait- and state-level confidence could help explain effects such as post-error slowing while being consistent with the findings of the present dissertation. The finding that greater prospective confidence is associated with faster drift rates is somewhat consistent with Bandura's (1997) theory of self-efficacy. Bandura suggests that when people have low self-efficacy, they tend to ruminate on what can go wrong, “depress themselves... and impair their level of functioning” (Bandura, 1997, p. 140). It is possible that participants with low confidence focused on potential negative outcomes, which impeded their ability to gather evidence. Indeed, attention has been shown to affect drift rates (Krajbich et al., 2010; Nunez et al., 2017). Interestingly, increasing drift rates and thresholds would both be expected to increase retrospective confidence judgments (Pleskac & Busemeyer, 2010). Therefore, prospective confidence could directly inform retrospective confidence judgments,

modulate decision processes in ways that increase or decrease retrospective confidence (e.g., motivating more cautious decision making), or both.

Research on confidence in judgment and decision making and cognitive psychology has largely focused on retrospective confidence. Moreover, cognitive modeling of decision making and confidence has largely focused on explaining intraindividual variance in retrospective confidence. However, our findings suggest that prospective confidence plays a significant role in modulating decision processes and in determining retrospective confidence—particularly confidence bias. Just as process models of decision making have been developed to consider attention (Krajbich et al., 2010), process models of decision making and confidence should expand to account for prospective confidence.

Affect and Confidence

We propose a mechanism of affect and confidence similar to the somatic marker hypothesis (Bechara et al., 1997; Bechara & Damasio, 2005; Damasio, 1994; Damasio et al., 1996): As people gain experience in a task, they have affective responses to the task based on their perceived performance (as informed by objective feedback and judgments of performance). Over time, affective responses are indexed and associated with the task. When a person faces a similar task, they experience anticipatory affect that informs their prospective confidence and guides their interaction with the task. It is also possible that anticipatory affect makes emotionally congruent past experiences more available to inform prospective confidence (Bower, 1981; Peters et al., 2006). Supporting the idea that affect may be involved in learning confidence and informing prospective confidence is that physiological responses associated with affective responses (i.e., electrodermal activity [EDA]) tend to fluctuate when students see a difficult

problem and receive feedback (Fritz et al., 2014; Horvers et al., 2021; Malmberg et al., 2019; Wampfler et al., 2019). We further hypothesize that when people make decisions, they may experience anticipatory affect—possibly automatically elicited by decision speed or experienced disfluency—that informs their retrospective confidence judgments. Indeed, we find that hindsight training muted affective responses and reduced calibration. Moreover, we observed that more negative affect toward negative feedback during training was associated with greater retrospective confidence sensitivity.

Affect may play a large part in the effects of prospective confidence on decision processes and retrospective confidence. More positive affect associated with greater prospective confidence may motivate people to invest more effort in their decisions. Consistently, Peters and colleagues (2003) found that affect motivated greater deliberation when valuating real versus hypothetical lottery tickets. Additionally, negative affect associated with lower prospective confidence might spotlight past negative experiences or cause people to ruminate on their doubts impeding the evidence accumulation process (Bandura, 1997; Bower, 1981; Peters et al., 2006). Affect might even underlie the relation between prospective and retrospective confidence. Voodla and colleagues (2024) believed that affective valence would respond to expectancy violations in the progress of gathering evidence to make perceptual decisions. Thus, they expected that when they cued a trial as difficult and it was easy, that it would make people feel more positive afterward than if it was cued as easy and was easy. Instead, they found that expected and actual trial ease had additive effects on affective valence after the trial. This pattern of results is consistent with our hypothesized mechanism that difficulty cues trigger an affective response and ultimately

inform retrospective confidence, similar to how difficulty cues inform retrospective confidence (Van Marcke et al., 2024; Boldt et al., 2019).

In the hindsight condition, participants' confidence appears to have been based more on their memory of their perceived accuracy and less on their affective experience compared to foresight-with-feedback participants. In Study 1, we observed a lower level of overall retrospective confidence following the test phase for participants in the hindsight condition than the foresight-with-feedback condition, despite having similar levels of prospective confidence. This lower retrospective confidence for hindsight participants appeared to carry over to lower prospective confidence after one week, while foresight-with-feedback participants maintained their level of prospective confidence. We believe that the affective experiences protected foresight-with-feedback participants from losing confidence because the affective responses made the feedback more memorable (Levine & Pizarro, 2004). Lacking strong affective responses, participants' memory for the training phase may not have held up against the experience of the test phase.

Conclusion

Overall, our results suggest that experiencing hindsight bias during training does have consequences for future decision making. However, the assumption that hindsight bias causes *overconfidence*—or even increases confidence—is not supported by the present studies (though we maintain that these are possible effects). Furthermore, our results suggest that prospective confidence plays a significant role in how people decide and their confidence in their decisions, retrospectively. These findings have important implications for how confidence and decision making dynamically influence each other over time. Lastly, we showed that affective responses to feedback predict how confidence changes with experience and may be important for people to

distinguish between good and bad decisions in their retrospective confidence judgments. We feel that prospective confidence as a construct is a promising frontier for the study of judgments and decisions. Prospective confidence sits at the intersection of social and cognitive psychology and is applicable in a wide range of domains. We hope that the findings and research questions generated from the present dissertation will inform future research on the interactions of confidence and decision making.

REFERENCES CITED

- Aguilar-Lleyda, D., Lemarchand, M., & de Gardelle, V. (2020). Confidence as a priority signal. *Psychological Science, 31*(9), 1084–1096. <https://doi.org/10.1177/0956797620925039>
- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. *The Self in Social Judgment, 1*, 85–106.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576.
- Ariely, D., Tung Au, W., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied, 6*, 130–147. <https://doi.org/10.1037/1076-898X.6.2.130>
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 37*(1), 93–110.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6, Pt.1), 359–372. <https://doi.org/10.1037/h0043445>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
https://doi.org/10.1007/SpringerReference_223312
- Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, *60*(6), 941–951. <https://doi.org/10.1037/0022-3514.60.6.941>
- Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929–945. <https://doi.org/10.1037/0096-1523.24.3.929>
- Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372.
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.
<https://doi.org/10.1126/science.275.5304.1293>
- Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520–1531. <https://doi.org/10.1037/xhp0000404>
- Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence predictions affect performance confidence and neural preparation in perceptual decision making. *Scientific Reports*, *9*(1), Article 1. <https://doi.org/10.1038/s41598-019-40681-9>
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148.
<https://doi.org/10.1037/0003-066X.36.2.129>

- Bradfield, A., & Wells, G. L. (2005). Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Memory & Cognition*, 33(1), 120–130. <https://doi.org/10.3758/BF03195302>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.
- Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, 45(1), 223–241. [https://doi.org/10.1016/0001-6918\(80\)90034-7](https://doi.org/10.1016/0001-6918(80)90034-7)
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- Brown, J. (1984). Effects of induced mood on causal attributions for success and failure. *Motivation and Emotion*, 8(4), 343–353. <https://doi.org/10.1007/BF00991872>
- Bruine de Bruin, W., Okan, Y., Krishnamurti, T., & Huffman, M. D. (2023). The role of confidence and knowledge in intentions to (not) seek care for hypertension: Evidence from a national survey. *Medical Decision Making*, 0272989X221148196. <https://doi.org/10.1177/0272989X221148196>
- Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, 31(3), 628–641.
- Carlebach, N., & Yeung, N. (2020). Subjective confidence acts as an internal cost-benefit factor when choosing between tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 46(7), 729–748. <https://doi.org/10.1037/xhp0000747>

- Castel, A. D., McCabe, D. P., & Roediger III, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*, 107–111.
<https://doi.org/10.3758/BF03194036>
- Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of the National Academy of Sciences*, *108*(supplement_3), 15655–15659. <https://doi.org/10.1073/pnas.1010658108>
- Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, *21*(9), 1751–1765. <https://doi.org/10.1162/jocn.2009.21123>
- Culot, C., Fantini-Hauwel, C., & Gevers, W. (2021). The influence of sad mood induction on task performance and metacognition. *Quarterly Journal of Experimental Psychology*, *74*(9), 1605–1614. <https://doi.org/10.1177/17470218211004205>
- Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Avon Books: New York.
- Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420.
- Dawson, N. V., Connors, A. F., Speroff, T., Kemka, A., Shaw, P., & Arkes, H. R. (1993). Hemodynamic assessment in managing the critically ill: Is physician confidence warranted? *Medical Decision Making*, *13*(3), 258–266.
<https://doi.org/10.1177/0272989X9301300314>

- de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science*, *25*(6), 1286–1288.
<https://doi.org/10.1177/0956797614528956>
- Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *eLife*, *8*, e43499. <https://doi.org/10.7554/eLife.43499>
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87.
- Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. <https://doi.org/10.3758/s13414-011-0243-2>
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*, 5–17.
<https://doi.org/10.1037/0022-3514.84.1.5>
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*(5), 395–416. <https://doi.org/10.1037/0033-295X.85.5.395>
- Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*(6), 1057–1070.
<https://doi.org/10.1016/j.neuron.2012.12.002>
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680.

- Feather, N. T. (1966). Effects of prior success and failure on expectations of success and subsequent performance. *Journal of Personality and Social Psychology*, 3, 287–298.
<https://doi.org/10.1037/h0022965>
- Fengler, A., Xu, Y., Bera, K., Omar, A., & Frank, M. J. (In Preparation). *HSSM: A generalized toolbox for hierarchical bayesian estimation of computational models in cognitive neuroscience*.
- Fieker, M., Moritz, S., Köther, U., & Jelinek, L. (2016). Emotion recognition in depression: An investigation of performance and response confidence in adult female patients with depression. *Psychiatry Research*, 242, 226–232.
<https://doi.org/10.1016/j.psychres.2016.05.037>
- Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13(1), 1–17.
[https://doi.org/10.1002/\(SICI\)1099-0771\(200001/03\)13:1<1::AID-BDM333>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S)
- Fischhoff, B. (1975). Hindsight \neq foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299.
- Fischhoff, B. (2007). An early history of hindsight research. *Social Cognition*, 25(1), 10–13.
- Fleck, M. S., Daselaar, S. M., Dobbins, I. G., & Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex*, 16(11), 1623–1630.
<https://doi.org/10.1093/cercor/bhj097>

- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338–1349.
<https://doi.org/10.1098/rstb.2011.0417>
- Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, 32(18), 6117–6125.
<https://doi.org/10.1523/JNEUROSCI.6489-11.2012>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00443>
- Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, 2016(1), niw018.
<https://doi.org/10.1093/nc/niw018>
- Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., & Züger, M. (2014). Using physiological measures to assess task difficulty in software development. *Proceedings of the 36th International Conference on Software Engineering*, 402–413.
<https://doi.org/10.1145/2568225.2568266>
- Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L., & Cleare, A. J. (2005). Depression, confidence, and decision: Evidence against depressive realism. *Journal of Psychopathology and Behavioral Assessment*, 27(4), 243–252. <https://doi.org/10.1007/s10862-005-2404-x>
- Hamaker, E. L. (2023). The within-between dispute in cross-lagged panel research and how to move forward. *Psychological Methods*. Advance online publication.
<https://doi.org/10.1037/met0000600>

- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327.
- Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 605–619.
- Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors*, *21*(23), 7869.
- Ifcher, J., & Zarghamee, H. (2014). Affect and overconfidence: A laboratory investigation. *Journal of Neuroscience, Psychology, and Economics*, *7*(3), 125-150.
<https://doi.org/10.1037/npe0000022>
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*(6), 401–405.
<https://doi.org/10.1111/j.1467-9280.1993.tb00589.x>
- Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science*, *29*(4), 521–536.
<https://doi.org/10.1177/0956797617740646>
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1–24. <https://doi.org/10.1006/jmla.1993.1001>
- Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*(2), 99–113. <https://doi.org/10.1177/1754073908100432>
- Koellinger, P., & Treffers, T. (2015). Joy leads to overconfidence, and a simple countermeasure. *PLOS ONE*, *10*(12), e0143263. <https://doi.org/10.1371/journal.pone.0143263>

- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816789.012>
- Koriat, A., & Adiv, S. (2012). Confidence in one's social beliefs: Implications for belief justification. *Consciousness and Cognition, 21*(4), 1599–1616. <https://doi.org/10.1016/j.concog.2012.08.008>
- Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). The Guilford Press.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1981). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118. <https://doi.org/10.1037/0278-7393.6.2.107>
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*, 478–492. <https://doi.org/10.1016/j.jml.2005.01.001>
- Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292–1298. <https://doi.org/10.1038/nn.2635>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

- Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167.
<https://doi.org/10.1038/nn.4064>
- Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition*, *22*(5), 530–554. <https://doi.org/10.1521/soco.22.5.530.50767>
- Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, *131*(6), 803–855.
<https://doi.org/10.1037/0033-2909.131.6.803>
- Malmberg, J., Haataja, E., Seppänen, T., & Järvelä, S. (2019). Are we together or not? The temporal interplay of monitoring, physiological arousal and physiological synchrony during a collaborative exam. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 467–490. <https://doi.org/10.1007/s11412-019-09311-4>
- Markowitz, D. M., Hancock, J. T., Bailenson, J. N., & Reeves, B. (2019). Psychological and physiological effects of applying self-control to the mobile phone. *PLOS ONE*, *14*(11), e0224464. <https://doi.org/10.1371/journal.pone.0224464>
- Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52. <https://doi.org/10.1111/j.1745-6916.2007.00028.x>
- McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making*, *17*(4), 281–295.
- Mellers, B., Fincher, K., Drummond, C., & Bigony, M. (2013). Chapter 1 - Surprise: A belief or an emotion? In V. S. C. Pammi & N. Srinivasan (Eds.), *Progress in Brain Research* (Vol. 202, pp. 3–19). Elsevier. <https://doi.org/10.1016/B978-0-444-62604-2.00001-0>

- Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817.
<https://doi.org/10.1287/mnsc.2021.4294>
- Modirrousta, M., & Fellows, L. K. (2008). Medial prefrontal cortex plays a critical and selective role in ‘feeling of knowing’ meta-memory judgments. *Neuropsychologia*, *46*(12), 2958–2965. <https://doi.org/10.1016/j.neuropsychologia.2008.06.011>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology*, *76*, 117–130. <https://doi.org/10.1016/j.jmp.2016.03.003>
- Pannu, J. K., Kaszniak, A. W., & Rapcsak, S. Z. (2005). Metamemory for faces following frontal lobe damage. *Journal of the International Neuropsychological Society*, *11*(6), 668–676.
<https://doi.org/10.1017/S1355617705050873>
- Parker, A. M., & Stone, E. R. (2014). Identifying the effects of unjustified confidence versus overconfidence: Lessons learned from two analytic methods. *Journal of Behavioral Decision Making*, *27*(2), 134–145.
- Pescetelli, N., Hauperich, A.-K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. *Cognition*, *215*, 104810.
<https://doi.org/10.1016/j.cognition.2021.104810>
- Peters, E., Lipkus, I., & Diefenbach, M. A. (2006). The functions of affect in health communications and in the construction of health preferences. *Journal of Communication*, *56*(suppl_1), S140–S162.

- Peters, E., & Shoots-Reinhard, B. (2022). Numeracy and the motivational mind: The power of numeric self-efficacy. *Medical Decision Making*, *42*(6), 729–740.
<https://doi.org/10.1177/0272989X221099904>
- Peters, E., Shoots-Reinhard, B., Evans, A. T., Shoben, A., Klein, E., Tompkins, M. K., Romer, D., & Tusler, M. (2019). Pictorial warning labels and memory for cigarette health-risk information over time. *Annals of Behavioral Medicine*, *53*(4), 358–371.
<https://doi.org/10.1093/abm/kay050>
- Peters, E., Slovic, P., & Gregory, R. (2003). The role of affect in the WTA/WTP disparity. *Journal of Behavioral Decision Making*, *16*(4), 309–330.
<https://doi.org/10.1002/bdm.448>
- Peters, E., Tompkins, M. K., Knoll, M. A. Z., Ardoin, S. P., Shoots-Reinhard, B., & Meara, A. S. (2019). Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes. *Proceedings of the National Academy of Sciences*, *116*(39), 19386–19391. <https://doi.org/10.1073/pnas.1903126116>
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901.
<https://doi.org/10.1037/a0019737>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.
- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127-140.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*, 90–138. [https://doi.org/10.1016/0010-0285\(87\)90005-3](https://doi.org/10.1016/0010-0285(87)90005-3)

- Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451. <https://doi.org/10.1037/0278-7393.18.3.435>
- Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426.
- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), Article 1. <https://doi.org/10.1038/s41467-019-09075-3>
- Rouault, M., Lebreton, M., & Pessiglione, M. (2023). A shared brain system forming confidence judgment across cognitive domains. *Cerebral Cortex*, *33*(4), 1426–1439. <https://doi.org/10.1093/cercor/bhac146>
- Salem-Garcia, N., Palminteri, S., & Lebreton, M. (2023). Linking confidence biases to reinforcement-learning processes. *Psychological Review*, *130*(4), 1017–1043. <https://doi.org/10.1037/rev0000424>
- Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28.
- Schnyer, D. M., Nicholls, L., & Verfaellie, M. (2005). The role of VMPC in metamemorial judgments of content retrievability. *Journal of Cognitive Neuroscience*, *17*(5), 832–846. <https://doi.org/10.1162/0898929053747694>
- Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, *42*(7), 957–966. <https://doi.org/10.1016/j.neuropsychologia.2003.11.020>

- Schraw, G. (2009). *Measuring metacognitive judgments*. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429).
Routledge/Taylor & Francis Group.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460–475.
- Shen, L., Fishbach, A., & Hsee, C. K. (2014). The motivating-uncertainty effect: Uncertainty increases resource investment in the process of reward pursuit. *Journal of Consumer Research, 41*(5), 1301–1315.
- Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science, 30*(4), 307–314.
<https://doi.org/10.1177/09637214211009510>
- Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*(1), 29–53.
- Silverstein, M. C., Bjälkebring, P., Shoots-Reinhard, B., & Peters, E. (2023). The numeric understanding measures: Developing and validating adaptive and nonadaptive numeracy scales. *Judgment and Decision Making, 18*, e19. <https://doi.org/10.1017/jdm.2023.17>
- Singer, M., & Tiede, H. L. (2008). Feeling of knowing and duration of unsuccessful memory search. *Memory & Cognition, 36*(3), 588–597. <https://doi.org/10.3758/MC.36.3.588>
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299–314.
- Son, L. K. (2004). Spacing one’s study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 601–604.
<https://doi.org/10.1037/0278-7393.30.3.601>

- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204–221.
<https://doi.org/10.1037/0278-7393.26.1.204>
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, *100*(4), 961–976.
- Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organizational Behavior and Human Decision Processes*, *59*(3), 452–474. <https://doi.org/10.1006/obhd.1994.1069>
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140.
<https://doi.org/10.1016/j.cogpsych.2011.06.001>
- van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence is the bridge between multi-stage decisions. *Current Biology*, *26*(23), 3157–3168. <https://doi.org/10.1016/j.cub.2016.10.021>
- Van Marcke, H., Denmat, P. L., Verguts, T., & Desender, K. (2024). Manipulating prior beliefs causally induces under- and overconfidence. *Psychological Science*, *35*(4), 358–375.
<https://doi.org/10.1177/09567976241231572>
- Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology*, *87*(3), 506–516. <https://doi.org/10.1037/0021-9010.87.3.506>
- Vickers, D., Smith, P., Burt, J., & Brown, M. (1985). Experimental paradigms emphasising state or process limitations: II effects on confidence. *Acta Psychologica*, *59*(2), 163–193.
[https://doi.org/10.1016/0001-6918\(85\)90018-6](https://doi.org/10.1016/0001-6918(85)90018-6)

- Voodla, A., Uusberg, A., & Desender, K. (2024). Affective valence does not reflect progress prediction errors in perceptual decisions. *Cognitive, Affective, & Behavioral Neuroscience, 24*(1), 60–71. <https://doi.org/10.3758/s13415-023-01147-8>
- Wampfler, R., Klingler, S., Solenthaler, B., Schinazi, V., & Gross, M. (2019). Affective state prediction in a mobile setting using wearable biometric sensors and stylus. *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019. International Educational Data Mining Society (IEDMS) 2019*, 198–207. <https://doi.org/10.3929/ethz-b-000393912>
- Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory, 11*(4–5), 411–419.
- Winman, A., Juslin, P., & Björkman, 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*(2), 415–431.
- Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K. H., Brown, J. W., Nelissen, N., & Rushworth, M. F. S. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature Communications, 7*(1), Article 1. <https://doi.org/10.1038/ncomms12327>
- Wright, J., & Mischel, W. (1982). Influence of affect on cognitive social learning person variables. *Journal of Personality and Social Psychology, 43*(5), 901–914. <https://doi.org/10.1037/0022-3514.43.5.901>
- Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance, 30*(1), 132–156. [https://doi.org/10.1016/0030-5073\(82\)90237-9](https://doi.org/10.1016/0030-5073(82)90237-9)

- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321. <https://doi.org/10.1098/rstb.2011.0416>
- Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., Horie, K., Sato, S., Kawashima, R., & Nakamura, K. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience Research*, 68(3), 199–206. <https://doi.org/10.1016/j.neures.2010.07.2041>
- Zacharopoulos, G., Binetti, N., Walsh, V., & Kanai, R. (2014). The effect of self-efficacy on visual discrimination sensitivity. *PLOS ONE*, 9(10), e109392. <https://doi.org/10.1371/journal.pone.0109392>
- Zhao, W. J., Coady, A., & Bhatia, S. (2022). Computational mechanisms for context-based behavioral interventions: A large-scale analysis. *Proceedings of the National Academy of Sciences*, 119(15), e2114914119. <https://doi.org/10.1073/pnas.2114914119>