PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

THE EVOLUTION OF OVERCONFIDENCE: A GROWTH CURVE MODELING STUDY OF OVERESTIMATION, ACCURACY AND CONFIDENCE.

FRANCISCO ANTONIO CARRASCO

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering (or Doctor in Engineering Sciences)

Advisor:
TOMÁS REYES

Santiago de Chile, (April, 2018)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

THE EVOLUTION OF OVERCONFIDENCE: A GROWTH CURVE MODELING STUDY OF OVERESTIMATION, ACCURACY AND CONFIDENCE.

FRANCISCO ANTONIO CARRASCO

Members of the Committee:

TOMÁS REYES

ALEJANDRO MAC CAWLEY

EDGAR KAUSEL

CHRISTIAN OBERLI

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering (o Doctor in Engineering Sciences)

Santiago de Chile, (April, 2018)

Gratefully to my parents, boyfriend, sister and nephew

AKNOWLEDGMENTS

First and foremost, I would like to thank my family for their support all over this years, constantly helping me to achieve my goals.

Second, Tomas Reyes, for his guidance during the whole time. He patienly taught me how to investigate and carry out a reaserch.

Third, I would like to thank the thesis committee, professors Christian Oberli, Alejandro Mac Cawley and specially Edgar Kausel for their important comments, willingness and insightful advice along this process.

CONTENTS

Page

AKNOWLEDGMENTS iv

CONTENTS v

LIST OF TABLES viii

ABSTRACT x

RESUMEN xi

1. ARTICLE BACKGROUND 1

1.1 Introduction 1

1.2 Theory and Hypotheses 2

1.2.1 Overconfidence Literature 2

1.2.2 The Importance: Management, CEOs 2

1.3 Methodology 3

1.3.1 The Experimental Design 3

1.3.2 Econometric Analyses 5

1.3 Main Results and Conclusion 6

2. THE EVOLUTION OF OVERCONFIDENCE: A GROWTH CURVE MODELING STUDYING OF SUBJECTIVE DIFFICULTY, ACCURACY AND CONFIDENCE. 7

2.1 Introduction 7

2.2 Experience and Overconfidence 8

2.3 Feedback and Overconfidence 10

2.3 Confidence and Accuracy 11

2.3 Overview of the studies 12
3. RESEARCH CASES 12
3.1 STUDY 1 12
3.1.1 Method 12
3.1.1.1 Participants, Design and Procedure 12
3.1.2 Variables 14
3.1.2.1 Experience. 14
Experience was operationalized simply as the trial number. Because the experiment included a total of 20 trials, experience ranged from 1 to 20. 14
3.1.2.2 Actual Performance. 14
3.1.2.3 Level of confidence. 14
3.1.2.4 Overconfidence. 15
3.1.3 Results and Discussion 15
3.2 STUDY 2 20
3.2.1 Method 20
3.2.2 Results and Discussion 21
3.3 STUDY 3 25
3.3.1 Method 25
3.3.2 Results and Discussion 26
3.4 STUDY 4 31
3.4.1 Method 31
3.4.1.1 Participants and Procedure 31
3.4.2 Results and Discussion 31
4. GENERAL DISCUSSION 35

4.1 General description of the main findings 35
4.2 Discrepancies across studies..36
4.3 Implications 37
4.4 Limitations..39
4.5 Conclusion..40
REFERENCES..41

LIST OF TABLES

Table 1: Study 1: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors. 16

Table 2: Study 2: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors 22

Table 4: Study 3: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors. For confidence and overconfidence, interaction between Experience and Framing is significant. The quadratic term of experience is also significant, supporting once again H3. 27

Table 6: Study 4: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors. 32

LIST OF FIGURES

Figure 1: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback)............... 18

Figure 2: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback)................................. 19

Figure 3: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback) 20

Figure 4: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback).......... 22

Figure 5: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback)............................ 23

Figure 6: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback)...................... 24

Figure 7: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (No Feedback)............. 28

Figure 8: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With feedback)............................. 29

Figure 9: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With Feedback)...................... 30

Figure 10: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 4 (With Feedback)*e* 32

Figure 11: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 4 (With feedback). 33

Figure 12: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With Feedback)...................... 34

ABSTRACT

Confidence and overconfidence have been widely studied in the judgment and decision making literature. However, previous research has paid little attention to how both constructs evolve and develop through time. In the present study, we examined how subjective difficulty and task framing, the experience of performing a task, and outcome feedback affect decision makers' confidence and overconfidence across trials.

We conducted a series of 4 studies, ran both in Latin America and the Center for Decision Research at the University of Chicago, involving a total of 325 participants. We manipulated the perceived difficulty of the task by telling participants about how others had performed (i.e., easy task vs. hard task). Participants were asked to estimate the amount of words in a text, as well as their confidence levels, in 20 trials. We examined the results using mixed effect regression models, growth curve modeling and lag operator analyses. We found that, when people were not provided with feedback, the framing manipulation affected confidence and overconfidence not only at the beginning of the experiment but also consistently across trials. Participants who were told the task was easy had increased confidence and overconfidence throughout the experiment, in contrast to what is found in the hard-easy effect (i.e., hard tasks lead to larger overconfidence). However, feedback moderated the effect of framing on confidence and overconfidence across trials, such that people's overconfidence when told the task was easy greatly diminished over time.

We make a contribution to the literature of overconfidence and the hard/easy effect. We conclude that both actual and perceived difficulty are important antecedents of overconfidence. We also show that feedback has a curvilinear effect on overconfidence, such that feedback in the first trials has stronger effect on overconfidence than feedback in the final trials.

Key words: Overconfidence; Perceived Difficulty; Feedback

RESUMEN

El exceso de confianza, más conocido como *overconfidence,* es uno de los sesgos cognitivos más comunes que afectan la toma de decisiones. Este sesgo es una sobreestimacion de las capacidades personales. Este sesgo se compone de dos variables que están siempre interactuando: exactitud y confianza. El exceso de confianza se genera cuando la confianza generada excede el nivel de exactitud o precision que tiene una persona a la hora de entregar un juicio o tomar una decisión.

Overconfidence ha sido ampliamente estudiado por la literatura sobre toma de decisiones y *management*. Se ha vinculado con emprendedores más arriesgados, con líderes cuyas empresas invierten más en investigación y desarrollo así como que suelen destruir valor en los procesos de fusiones y adquisiciones corporativas.

Ahora bien, a lo largo de esta investigación se analizará el exceso de confianza como un fenomeno dinámico, variable a lo largo del tiempo. Esto se diferencia de la literatura actual que toma este sesgo como constante e inherente a la persona. Primero se analiza cuál es el efecto que tiene la dificultad percibida de una tarea en el exceso de confianza y como este efecto cambia a medida que se adquiere experiencia sobre la tarea. Luego de esto, se analiza como la introducción de *feedback* es capaz de reducir el exceso de confianza y cómo interactua con confianza y precisión para lograr este objetivo.

Los resultados obtenidos son que la experiencia por si sola es poco capaz para lograr un cambio significativo en el exceso de confianza. Por el contrario, *feeback* es altamente eficaz en calibrar a las personas, es decir, en eliminar el efecto del exceso de confianza y equiparar la confianza con la precisión.

1. ARTICLE BACKGROUND

1.1 Introduction

Overconfidence has been widely studied as a bias in the decision making process and has been pointed out as one of its most influential factors (Kahneman, 2011). Overconfidence is an overestimation of one's own performance, ability or accuracy of prediction. Hence, this bias is built by constant interaction between confidence and accuracy. This study will show how overconfidence changes when feedback is added to this interaction.

Overconfidence influences decisions in many different contexts. In entrepreneurship overconfidence has been positively related with new business creation because new business owners are naturally prone to have high expectations of success (Bernardo & Welch, 1997; Forbes, 2005; Minniti & Minniti, 2005). Career choice is also affected by overconfidence, Schulz & Thöni (2016) found there are systematic differences between confidence among different fields of study and that these differences may have implications for future earnings and the careers students select. Finally, personnel selection can also be affected by overconfidence. People who are presented with interview information exhibit more overconfidence than those who are presented only with scores (Kausel, Culbertson, & Madrid, 2016).

Although, the literature has vastly studied the effect of overconfidence and situations in which people are more prone to be overconfident, few articles have been written about the evolution of overconfidence when individuals are faced with new information, expectations, feedback or experience.

The remainder of this chapter is organized as follows. Section 1.2 discusses the main objectives of this research. Section 1.3 presents a literature review on overconfidence. Section 1.4 presents the methodology and data used. Finally, Section 1.5 exhibit the findings and conclusions of this study.

1.2 Theory and Hypotheses

1.2.1 Overconfidence Literature

It is a natural question to ask ourselves: why are people overconfident? (Anderson, Brion, Moore & Kennedy (2012). If it is useful to realistically compare one's abilities to others (e.g., Alicke, 1985; Dunning et al., 2004; Larrick, Burson, & Soll, 2007), then why are people still so prone to be overconfident? Anderson, Brion, Moore & Kennedy (2012) gathered three points of view given by scholars. The first on says that people tend to feel overconfident because confidence provides them with psychological satisfaction (Dunning, Leuenberger, & Sherman, 1995; Kunda, 1987). A second explanation points to cognitive process, as the one that produces mistakes in the assessment of one's own capabilities. For example, Kruger & Dunning (1999) think people luck the competence to understand their own incompetence. A third explanation claims that overconfidence provides social benefits. Studies show that social success may be increased by a biased self-perception (Alexander, 1987; Krebs & Denton, 1997; Leary, 2007; Trivers, 1985; von Hippel & Trivers, 2011; Waldman, 1994).

1.2.2 The Importance: Management, CEOs

From the many biases affecting judgmental decisions we focus on overconfidence. Overconfidence is particularly important in management, and its effects on CEOs have been widely discussed in the literature (Anderson et al., 2012, Malmendier and Tate, 2005, 2008; Busenitz and Barney, 1997; Zacharakis and Shepherd, 2001; Ho et al., 2016). This is mainly because of the importance of CEO's decisions and their effects on the organizations they lead. CEO's have been related to speculative bubbles (Scheinkman & Xiong, 2003), innovation (Forbes, 2005), failed mergers and acquisitions (Roll, 1986), recessions (Ho et al., 2016) and even great corporative failures such as Enron, Crossing and the National Kidney Foundation (Picone, Dagnino, and Minà, 2014). The problem with overconfident CEOs relies on their poor capability to make accurate forecasts about the future (Hribar and Yang, 2015). According to Hribar and Yang (2015), overconfident managers tend to overestimate future returns, making them poor decisions makers.

High self-confidence also builds a sense of superiority towards CEO’s subordinates. They believe that they have better than average abilities. And in the relation with their work partners, overconfident managers feel they have better skills to make decisions than them (Ferris, Jayaraman, and Sabherwal, 2013). This has to do with the poor capacity of them to hear feedback from others (Chen, Crossland, and Luo, 2015).

1.3 Methodology

The main objective of this study is to see how overconfidence evolves through time. In other words, we are taking overconfidence as a dynamic phenomenon that can change with experience and feedback.

1.3.1 The Experimental Design

We designed an online experiment. Each participant received a link through his or her personal email. We manipulated the perceived difficulty of a task, participants were randomly assigned into two different conditions: *hard* and *easy*. Those in the hard (easy) condition were told that: (a) the task was hard (easy); (b) people who had previously performed this task had poor (good) performance; (c) taking into account the way they were recruited, we expected they would have a poorer (better) performance than the average participant; and (d) people who had previously performed this task had taken more (less) time than the one we would give them.

After manipulating perceived difficulty to make the initial framing, we asked participants to estimate the number of words of 20 different excerpts of a Spanish version of The Little Prince. Each excerpt was unique and had between 200 to 400 words. Each one was shown for 8 seconds in the screen. After this, participants were unable to see the excerpt and were asked to type their estimation of the number of words. After each word estimation, their level of confidence was measured in order to study the influence of perceived difficulty and feedback in level of confidence and performance across different levels of experience and item difficulty.

We used four variables in the analyses: experience, accuracy, confidence and overconfidence. First, experience was operationalized as the number of excerpts (which we will call 'trials') the subject had already read and estimated. Because the experiment included a total of 20 trials, experience ranged from 1 to 20. While they performed it, it is likely that participants had a better sense of what the task was like, as well as its difficulty. Second, actual performance or accuracy was measured by using the participants' estimation of the number of words they believed the excerpt had. Actual performance was a binary variable at the trial level: If the participants' estimation was at most 50 words different from the actual number of words the excerpt had it was considered a correct answer. Otherwise, it was incorrect. Participants were told at the outset about how we would compute actual performance. Third, each participant was asked about his or her level of confidence after they estimated the amount of words in all 20 passages. For this, participants were asked to answer a 0 to 100% certainty measure. This measure was selected by dragging an option bar that was by default set at 50% confidence and it could be moved to the right or to the left in order to get to the desired level of confidence. An image of the option bar can be seen in the appendix. This allows us to understand how subjects react to perceived difficulty while feedback and experience are added in time. Finally overconfidence was operationalized as overestimation: the difference between confidence and accuracy. Given that this difference score tends to produce large standard errors, below we describe how we refined this measure based on a procedure suggested by Budescu and Johnson (2011).

We separated our research in four different studies. Studies 1, 2 and 4 were held out with students who were completing an engineering or management degree at a large university located in Santiago, Chile. Participants were offered course credits for participating. Study 3 was held in the Center for Decision Research at the University of Chicago Booth School of Business.

The first Study had the goal to test if experience and framing a task as difficult (vs. easy) have an effect on overconfidence over time. Study 1 is the only one that has no feedback. In study 2, 3 and 4 we provided participants with feedback about their performance after each estimation. More specifically, after they had already answered we showed them a message with the actual amount of words in the excerpt. In addition, we showed them the interval of words in which an answer was considered correct.

Feedback has been shown to be a useful strategy to reduce overconfidence, it is unclear whether the effect of feedback is linear over time. In other words, it is unclear whether it occurs slowly and linearly or it has an immediate, curvilinear effect. We do know that feedback has a curvilinear effect on accuracy, which is important to determine overconfidence. For example, Lagnado & Newell (2006) showed in a weather prediction task that participants improved accuracy after receiving feedback; however, this effect decreased over time and had a significant quadratic form. Consistent with this result, Mirman et al. (2009) examined the link between statistical segmentation and word learning and showed a fast learning at the beginning followed by an asymptotic plateau. These studies focused on the effect of feedback on accuracy only, but we believed there would be a similar effect of feedback on overconfidence.

The first objective of adding feedback to the experiment is to find out if it will have a curvilinear effect on overconfidence over time? In other words, will feedback have a stronger impact in earlier trials than in the last ones? The second objective is to find out if feedback moderates the effect of the perceived difficulty framing on time.

1.3.2 Econometric Analyses

To test our hypotheses and research questions we conducted multi-level regression analyses (Strohminger, Lewis, & Meyer, 2011; Waldman & Yammarino, 1999). We decided to use this type of modeling because of the hierarchical structure of the data. The experiment utilizes different framing conditions between participants—a level 2 variable. In contrast, experience varied within participants, and thus it was a level 1-variable. Because of this, we conducted regression analyses that allowed for random intercepts representing different levels of confidence and accuracy for different individuals. We conducted three different regression analyses for the three main dependent variables: accuracy, confidence, and overconfidence.

The focus of this investigation is to study how overconfidence changes over time. To do this, we want to answer more specific research questions. First, we examined if getting more experienced with a task implied a decrease in overconfidence. Second, we studied if experience reduced the effect of framing the perceived difficulty of a task on confidence, accuracy and overconfidence. In sum, we examined whether feedback decreased overconfidence over time. Finally, we tested whether accuracy at time *t* had lagged effects on confidence at time *t +1*. To test these ideas, we conducted four laboratory experiments across two different samples using a relatively simple, word-counting task.

1.3 Main Results and Conclusion

Results suggested that experience by itself is not effective in the reduction of overconfidence (Study 1). This effectiveness of experience depends on the beliefs people held regarding the task—how the task was framed: easy of hard. Those told the task would be easy decreased (slightly) their large initial overconfidence over time. Those told the task would be hard reduced their overconfidence in the initial trials; after 10 trials, their overconfidence increased, perhaps due to a fatigue (or discouragement) effect. Accuracy, more so than confidence, tended to explain these results.

Feedback, on the other hand, decreases overconfidence over time in a very effective way (Studies 2–4). This is particularly strong during the first trials; the effect faded in subsequent trials. Indeed, by trial 15, feedback does not reduce overconfidence. This could have two explanations. First, feedback may have been informative only at the outset. Stated differently, feedback is a very good hint for participants' estimations; after this, this information could be redundant. Second, this could be the result of a floor effect. For example, participants in Studies 3 and 4, by trial 10, were already presenting almost perfect calibration (no overconfidence). As such, there was no much room for improvement.

Finally, cross-lagged analyses showed that, among those who received feedback, accuracy at time t had an effect on confidence at $t + 1$. People who had a correct estimate were more likely to increase their confidence in a subsequent trial. (We should also note that accuracy had no effect on confidence).

2. THE EVOLUTION OF OVERCONFIDENCE: A GROWTH CURVE MODELING STUDYING OF SUBJECTIVE DIFFICULTY, ACCURACY AND CONFIDENCE.

2.1 Introduction

Overconfidence is an unwarranted faith in one's knowledge and abilities (Harvey, 1997; Moore & Healy, 2008; Pompian, 2006), which leads individuals to overestimate their performance and the accuracy of their predictions[1] (Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999). Some authors have argued that it is likely the most common decision bias (Kahneman, 2011), and is relevant in many contexts, from entrepreneurship (Bernardo & Welch, 2001; Forbes, 2005; Arenius & Minniti, 2005) to career choice (Schulz & Thöni, 2016) to personnel selection (Kausel, Culbertson, & Madrid, 2016). In short, overconfidence is widespread and has important consequences (but see Mandel & Barnes, 2014, 2018).

[1] We focus on what Moore and Healy (2008) term 'overestimation,' which is typically operationalized as the difference between confidence and actual performance (or accuracy).

Interestingly, previous research has focused on overconfidence as a fixed bias and has ignored the dynamic influence of experience or feedback over time (for two exceptions, see Pulford & Colman, 1997; Sanchez & Dunning (2018). In many environments, decision makers may experience various degrees of confidence regarding an outcome. For example, Berger and Pope (2011) studied basketball team outcomes when teams were either winning or losing by a small margin. They found that teams that were winning by one point at the end of the first half of a game were less likely to win the game than their opponents. This finding can be explained by overconfidence (see Rogers & Moore, 2015). Capturing an early lead leads to some degree of overconfidence, which in turn undermines effort. Importantly, note that the initially winning team was not necessarily overconfident at the beginning of the game. They likely *became* overconfident after the first half. This shows that overconfidence is a dynamic phenomenon, as opposed to stable or trait-like.

As such, in this paper, our main interest is to understand how overconfidence evolves over time. Our research is focused on four related issues. First, building on a previous study (Kausel, Pérez-Cotapo, Zhang, & Reyes, 2018), we analyzed how framing a task in terms of its perceived difficulty affects overconfidence over time. More specifically, we examined whether manipulating the perceived difficulty of a task has an effect only at the outset of the study (i.e., in the first trials of the experiment) or whether the effect remains consistent over trials. Second, we examined how individuals' levels of confidence and overconfidence evolve as they become more experienced with a task. Third, we analyzed whether feedback helps reduce overconfidence over time. Relatedly, assuming feedback does reduce overconfidence, we also studied whether the effect is incremental (constant across trials) or exponential (stronger in the first trials than in the final trials). Finally, we examined whether accuracy has a lagged effect on confidence.

2.2 Experience and Overconfidence

In a series of experiments, Kausel et al. (2018) found that framing a task as easy or difficult has an effect on overconfidence. Participants were told that an upcoming task was either hard or easy and this expectation affected their level of confidence and, as a result, their overconfidence. A critical driver of this result was subjective difficulty: Participants who were told that the task was hard had lower levels of confidence and overconfidence in their performance despite unchanged accuracy, while participants who were told the task was easy had greater levels of confidence and overconfidence.

Our interest in this study is understanding how overconfidence develops over time. We sought to replicate, but also extend, Kausel et al.'s (2018) findings by examining whether the perceived difficulty manipulation has an effect when individuals have experience with the task. This is an important question because, as Hertwig, Barron, Weber, and Erev (2004) stated, "decisions from experience and decisions from description can lead to dramatically different choice behavior" (p. 534). For example, Hertwig et al. (2004; see also Hertwig & Erev, 2009) proposed that individuals tend to overestimate rare events when they learn about these events from a written description. In contrast, this pattern disappears when individuals have experience with a task. In fact, researchers have proposed that decision makers are less likely to fall prey to cognitive biases when they have more experience with a task (Feng & Seasholes, 2005).

Previous research seems to back this claim. For example, experience among financial analysts and professionals has been found to reduce their levels of overconfidence (e.g., Odean, 1998; Gloede & Menkhoff, 2011). Similarly, Hansson, Juslin, and Winman (2008) found that task experience reduced overconfidence in an experimental task where participants estimated a company's income. Pulford & Colman (1997) using almanac questions and a repeated-measures design, found that participants showed less overconfidence in the final set of questions than in earlier sets. However, other studies have found that more experience tends to *increase* decision makers' overconfidence. Even within the financial literature we mentioned above, Glaser and Weber (2007; see also Mishra & Metilda, 2015) found investment experience to be positively related to overconfidence. Menkhoff, Schmeling, and Schmidt (2013), in an online experiment, also found that both investment experience and age were related to overconfidence. Deaves, Lüders, and Schröder (2010) used a theoretical model in which experienced stock market forecasters learned to be overconfident. Consistent with this model, they found that success typically leads to more overconfidence, whereas failure leads to less overconfidence. A recent study using a multi-cue learning task by Sanchez and Dunning (2018) suggested that the relationship between overconfidence and experience has a cubic form, meaning that participants exhibited no overconfidence at the beginning of the task, increased their overconfidence in early trials, and ended with leveled-off overconfidence levels.

Upon review of the literature, we noted a lack of consistent findings and diversity in study areas. Most papers exploring overconfidence and task experience are related to financial decisions, leaving other avenues unexplored. As a result, we ask the following research questions:

Research Question 1: Does framing a task as difficult (vs. easy) have an effect on overconfidence over time?

Research Question 2: Does experience with a task have an effect on overconfidence?

2.3 Feedback and Overconfidence

In a chapter published in S. Armstrong's (2001) *Principles of Forecasting*, Arkes (2001) proposed the following principles to avoid falling prey to overconfidence: "Make an explicit prediction and then obtain feedback…[and] conduct experiments to test prediction strategies" (pp. 504-507; see also Slaughter & Kausel, 2013). Indeed, research has shown that weather forecasters, experienced bridge players, and horseracing bettors tend to show excellent calibration (i.e., low to zero overconfidence) because they obtain immediate and unambiguous feedback (e.g., Stone & Opel, 2000; Tetlock, 2017; Pulford & Colman, 1997).

While evidence has shown that feedback is a useful strategy for reducing overconfidence, it is unclear whether its effect is linear over time. In other words, it is unclear whether the effect occurs slowly and linearly or if feedback has an abrupt effect on overconfidence. However, we know that feedback generally has a curvilinear effect on accuracy, which is an important indicator of overconfidence. For example, Lagnado and Newell (2006) showed in a weather prediction task that participants improved their accuracy after receiving feedback; however, this effect decreased over time and had a significant quadratic form. Consistent with this result, Mirman et al. (2009) examined the link between statistical segmentation and word learning, and showed fast learning at the beginning followed by an asymptotic plateau. These studies focused only on the effect of feedback on accuracy, but we anticipated that there would be a similar effect of feedback on overconfidence, because overconfidence (overestimation) is *defined* as confidence minus accuracy. Based on these findings, we hypothesized that the effect of feedback would be stronger in the first trials but weaker in later trials. We frame this in terms of the following research questions:

Research Question 3: Will feedback have a curvilinear effect on overconfidence over time, such that the effect is stronger impact in earlier trials of the task and weaker in later trials of the task?

Research Question 4: Will feedback moderate the effect of difficulty framing on overconfidence over time, such that the framing effect will be strong in earlier trials of the task, and will tend to disappear in later trials of the task?

2.3 Confidence and Accuracy

We are also interested in how confidence and accuracy are intertwined. More specifically, we sought to examine whether accuracy affected confidence in later trials. Deaves, Lüders, and Schröder (2010) used stock market forecasts in order to analyze how recent successful forecasts affected participants' confidence intervals in the following period. As expected, recent "hits" caused a narrowing in the confidence interval, resulting in increased levels of confidence and overconfidence. The authors also suggested that a higher market return generates higher market overconfidence; however, the impact of these events was limited. Thus, we hypothesized:

Hypothesis 1: Under feedback, there will be a lagged effect of accuracy on confidence in later trials, such that people who are accurate at time 1 will be more confident at t+1.

2.3 Overview of the studies

We used four studies to test our research questions and hypothesis. All of the studies utilized the same task and involved 20 trials. The task consisted of estimating the number of words in a text after seeing it for 8 seconds. Participants were randomly assigned into two framing conditions in all studies. Half of the participants were told the upcoming task was easy and the other half were told the task was difficult. Three factors differed across studies: feedback received by participants, location, and reward. First, participants received feedback in Studies 2, 3, and 4, but not in Study 1. Second, Studies 1, 2, and 4 were conducted at a Chilean university whereas Study 3 was conducted at the University of Chicago. Third, there were differences in the reward offered across studies. We describe the details below.

3. RESEARCH CASES

3.1 STUDY 1

3.1.1 Method

3.1.1.1 Participants, Design and Procedure

Studies 1, 2, and 4 were carried out in accordance with the recommendations of the Ethical Committee of Social Sciences, Art, and Humanities, from Pontificia Universidad Católica de Chile.

Participants included seventy-six working professionals who were completing a management degree at the Pontificia Universidad Católica de Chile located in Santiago, Chile. The mean age of participants was 38 years old (*SD* = 7.7) and 40% were female. Participants were offered course credit for participating and the opportunity to win a cash prize based on their performance. The experiment lasted approximately 15 minutes.

The participants were told that their performance would be assessed using a Brier (1950) score, which would be calculated based on both the accuracy of their estimates and their expressed confidence levels. We also provided examples to ensure the participants understood how their scores would be calculated. The participant with the lowest Brier score—assuming they had over 12 correct estimates—would earn the equivalent of a US$75 cash prize. We sought to reward good calibration with the offer of this cash prize.

The experiment was conducted online and each participant received the link to the experiment through their personal email. The main between-subjects independent variable was the framing difficulty manipulation. Participants were randomly assigned into two different conditions: *hard* or *easy*. Those in the hard (easy) condition were told that: (a) the task was hard (easy); (b) participants who had previously performed this task had poor (good) performance scores; (c) taking into account the way they were recruited and considering their profile, we expected them to have less (more) ability and thus poorer (better) performance than the average participant; and (d) participants who had previously performed this task had taken more (less) time than the current participants would be allowed.

After manipulating the perceived difficulty of the task in this way, we asked participants to perform a computer-based, 20-trial task with varying levels of difficulty. We elected a task in which people likely had no prior experience. Thus, participants did not have crystallized expectations about their task performance.

Participants received an email with a brief explanation of the experimental procedure but without details of the task. The email contained a link to the website with the experiment. Emails were sent individually to each participant, but with a previous random assignment to the conditions. Since the experiment was available online, participants could complete the task from anywhere with a personal computer and internet access.

The goal of the task was to estimate the number of words contained in different text passages, which were excerpts from the Spanish version of “The Little Prince.” Each text was shown to the participants for 8 seconds. After this, the passage disappeared and participants were asked to type in their word count estimates. Participants were then asked their levels of confidence in their estimates.

3.1.2 Variables

3.1.2.1 Experience.

Experience was operationalized simply as the trial number. Because the experiment included a total of 20 trials, experience ranged from 1 to 20.

3.1.2.2 Actual Performance.

Accuracy or actual performance was measured using the participants’ estimates of the number of words contained in the text. Accuracy was a binary variable at the trial level. If a participant’s estimate was within 50 words of the actual number of words the excerpt contained, the estimate was considered correct; otherwise, it was considered incorrect. We told participants how we would measure actual performance at the beginning of the study. Below we describe how we modeled and refined this measure.

3.1.2.3 Level of confidence.

Each participant was asked about their level of confidence after they estimated the number of words in each of the 20 text passages. Specifically, participants were asked the following question: “How confident are you that your estimate is correct?” The scale ranged from 0% to 100%. Participants could drag an option bar that was by default set at 50%. The bar could be moved to the right or to the left in order to reach the desired level of confidence.

3.1.2.4 Overconfidence.

We operationalized overconfidence as overestimation, which is the difference between confidence and accuracy. Because measuring overconfidence as this difference tends to produce large standard errors, we describe below how we refined this measure based on a procedure suggested by Budescu and Johnson (2011).

3.1.3 Results and Discussion

Our general analytical framework adopted a model-based approach. Budescu and Johnson (2011) proposed such an approach based on generalized linear mixed models for binary outcomes (Guo & Zhao, 2000) to study overconfidence and calibration (see also Mandel & Barnes, 2018). Accuracy could be modeled using a binary logistic link function with Framing, Experience, Experience^2, and Experience^2 x Framing (i.e., four terms) as predictors. Overconfidence was then operationalized as the discrepancy between confidence and predicted accuracy for each estimate.

To test our research questions and hypothesis, we used growth curve modeling (Bollen, 2007; Curran, Obeidat, & Losardo, 2010). In particular, we conducted multi-level regression analyses (Hox, 1998). We used this type of modeling because of the hierarchical structure of the data. The experiments used different framing conditions *between* participants, and experience varied *within* participants. Because of this, we conducted regression analyses that allowed for random intercepts and slopes representing different levels of confidence and accuracy for different individuals. We conducted three different multilevel regression analyses for the three main dependent variables: accuracy (logistic regression), confidence, and overconfidence.

Table 1 shows the results of these multilevel regression analyses. The left panel of Table 1 shows that the hard and easy framing conditions interacted with experience. Figure 1 shows that individuals who were told that the upcoming task would be easy decreased their overconfidence slightly but steadily over time (simple slope, b = -.004, p < 0.01). In contrast, for those told the task would be difficult, their overconfidence had a U-shape (the coefficient of the quadratic term for this condition was b = .001, p < .001). The manipulation decreased participants' overconfidence at an increasing rate up to an inflection point, which was around trial 10. As such, the answer to our Research Questions 1 and 2, regarding whether the framing manipulation (Research Question 1) and experience (Research Question 2) had an effect over time, is 'yes'. However, this effect is rather complex, especially with regard to how the framing manipulation interacts with experience.

Table 1

Study 1: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors.

	Overconfidence	Accuracy	Confidence
Hard Framing	0.061	-0.434	-0.029
Experience (Trial)	0.003	0.0170	0.007*
Hard Framing x Experience	-0.036***	0.132	-0.009
Experience x Experience	-0.000*	0.001	-0.000
Hard Framing x Experience x Experience	0.002***	-0.008*	0.000
Constant	0.300***	-0.873**	0.605***
N	1493	1493	1493
AIC	-3082.0	1837.4	-1619.8
BIC	-3029.6	1879.8	-1572.0
ICC	0.812	0.214	0.693

Note. For framing, hard = 1, easy = 0.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

AIC: Akaike information criteria

BIC: Bayesian information criterion

ICC: Interclass correlation

In order to better understand the effect of the framing manipulation over time, we examined the results of the multilevel regression using accuracy and confidence as dependent variables. In other words, we sought to understand whether the contrasting shapes of the overconfidence trend across trials between the different groups (those told the task would be easy vs. hard) was better explained by accuracy or confidence. The center panel of Table 1 shows the initial model using accuracy as the outcome, based on the binary logistic link function described above. It can be observed that the Experience^2 x Framing term has a significant effect on overconfidence.

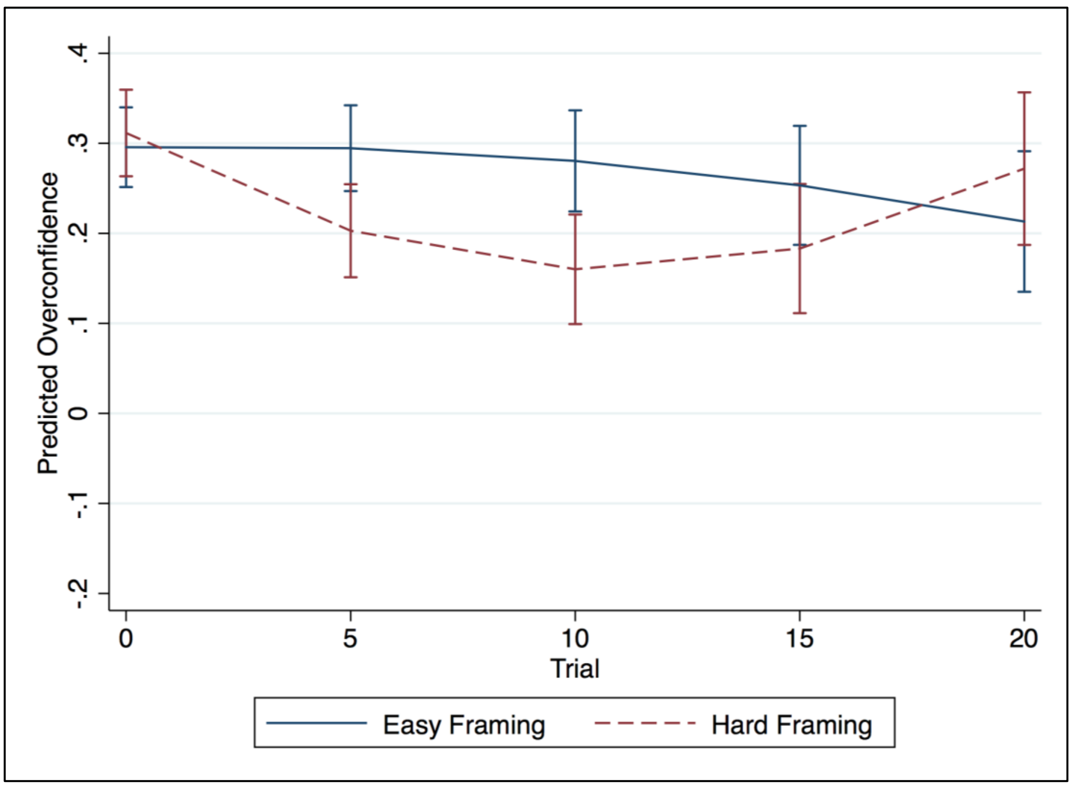

Figure 1: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback)

Figure 2 shows that those participants who were told the task would be easy tended to slightly increase their accuracy across trials. Participants who were told the task would be hard increased their accuracy in the initial trials, but there appeared to be a fatigue effect that led to a decrease in accuracy in the final trials. This is consistent with the idea that perceiving a task as challenging tends to increase participants' effort (Rogers & Moore, 2015); however, this also may lead to a fatigue or "discouragement" effect in the longer run. This figure also shows a steady improvement in accuracy for participants in the easy framing condition across the 20 trials. This is consistent with a simple slope analysis of this group across time (b = .042, $p < .001$). For participants in the hard framing condition, accuracy also increased, but only until trial 10; after this point, accuracy decreased. The quadratic term for this group was significant (b = -.007, $p < .05$). We believe that this could be explained by the fatigue (or 'discouragement') effect mentioned above.

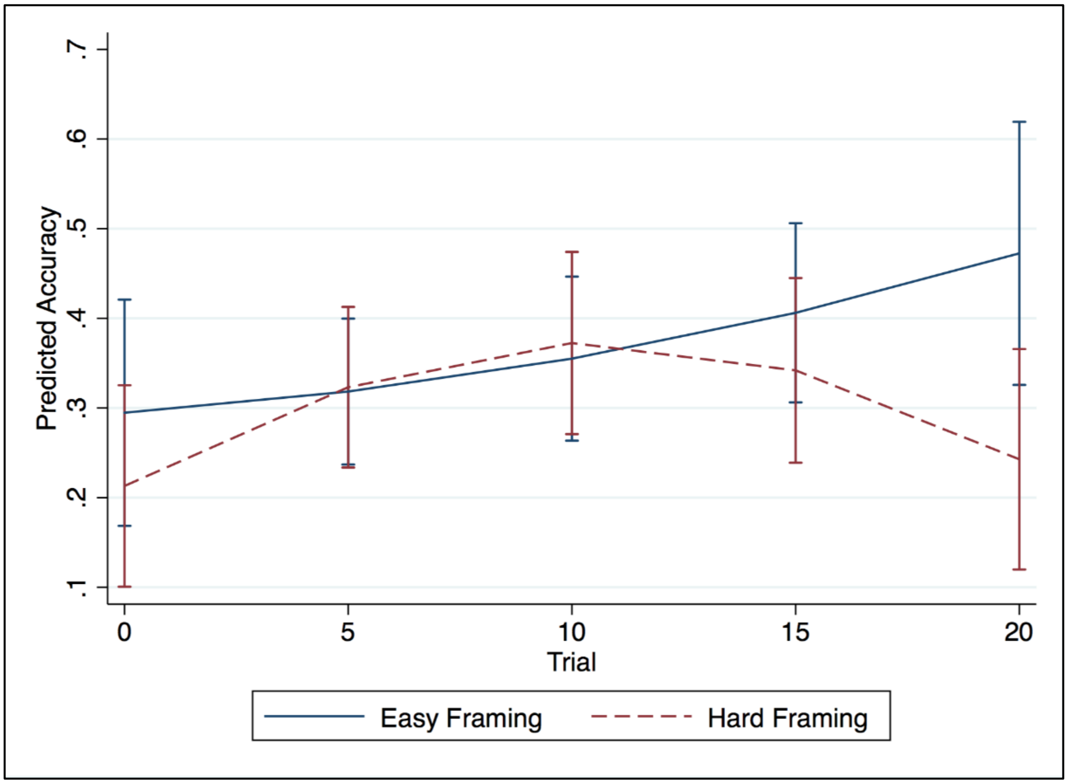

Figure 2: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback).

We then focused on confidence. As seen in the right panel of Table 1, the only significant term for confidence was trial (the proxy for experience). The relationship between confidence and experience was small but positive for participants in the easy framing condition (b = .004, $p < .001$); however, the relationship between confidence and experience was not significant for participants in the hard framing condition (b = .000, $p = .69$). Figure 3 shows the evolution of confidence across time.

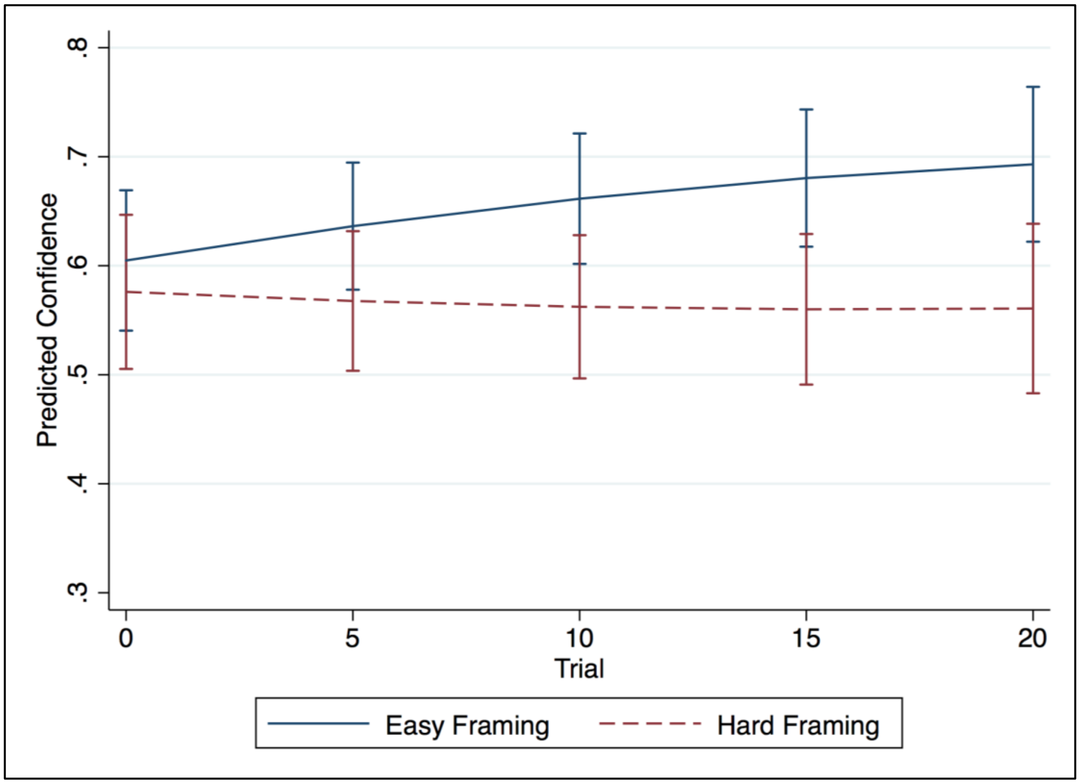

Figure 3: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 1 (no feedback)

In sum, overconfidence decreased linearly across trials for participants told the task would be easy, but had a U-shape for participants told the task would be hard. This seems to be explained mostly by accuracy. A potential explanation is a fatigue effect for participants in the hard framing condition, likely because they exerted more effort in the initial trials.

3.2 STUDY 2

3.2.1 Method

Participants included eighty-four working professionals who were completing a management degree at the Pontificia Universidad Católica de Chile. The mean age of the participants was 35 years old (*SD* = 7.6) and 37% were female. Participants were offered course credit for participating and the possibility of winning extra credit and US$75 based on the accuracy of their predictions. The experiment lasted approximately 15 minutes.

The procedure of Study 2 was very similar to that of Study 1. However, after each estimate, we provided participants with feedback about their performance. More specifically, we told participants the actual number of words contained in the text. In addition, we showed participants the range of answers that were considered correct.

3.2.2 Results and Discussion

Table 2 shows the multilevel regression analyses for overconfidence, accuracy (binary logistic link function), and confidence. The answer to Research Question 3, which asked whether feedback would have a curvilinear effect on overconfidence, is 'yes.' Research Question 4 asked whether feedback moderates the effect of framing on overconfidence across time. The quadratic interaction term Framing x Experience^2 was not significant. Thus, the answer to Research Question 4 is 'no.'

	Overconfidence	Accuracy	Confidence
Hard Framing	-0.033	0.105	-0.009
Experience (Trial)	-0.020***	0.160**	0.018***
Hard Framing x Experience	0.006*	-0.066	-0.009*
Experience x Experience	0.001***	-0.006*	-0.001***
Hard Framing x Experience x Experience	-0.000	0.003	0.000*
Constant	0.219***	-0.433	0.614***
N	1658	1658	1658
AIC	-3251.4	2223.4	-2126.7
BIC	-3198.2	2266.8	-2077.9
ICC	0.794	0.0285	0.689

Note. For framing, hard = 1, easy = 0.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2

Study 2: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors.

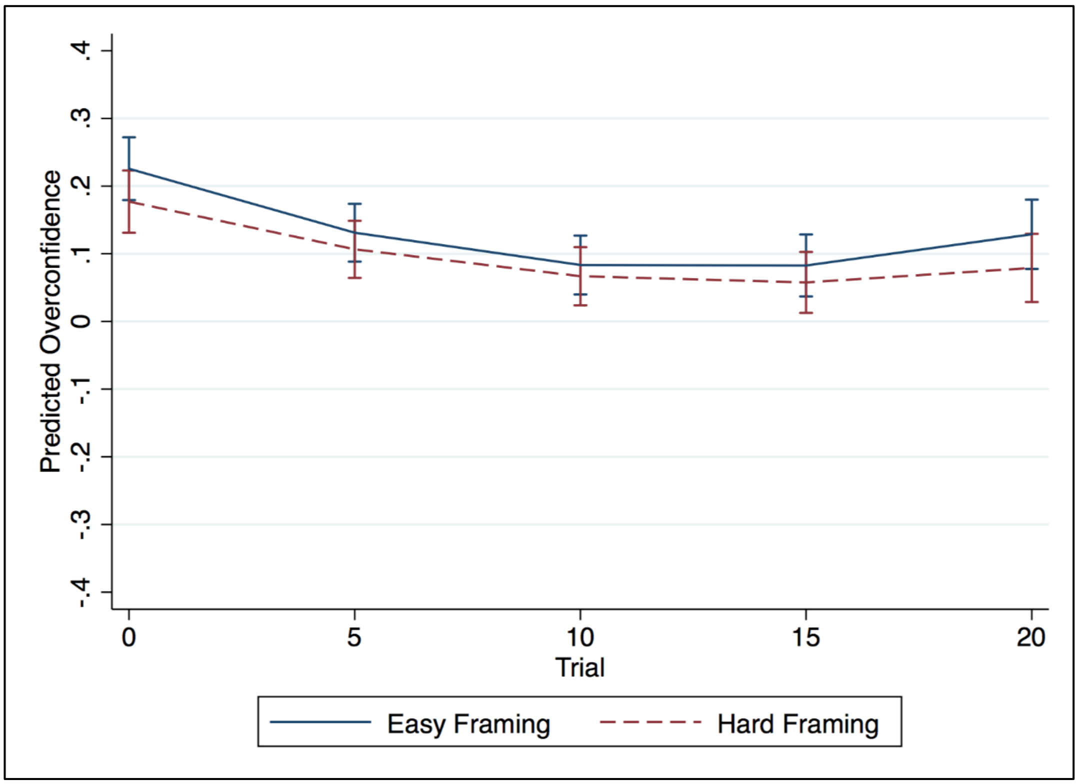

Figure 4: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback)

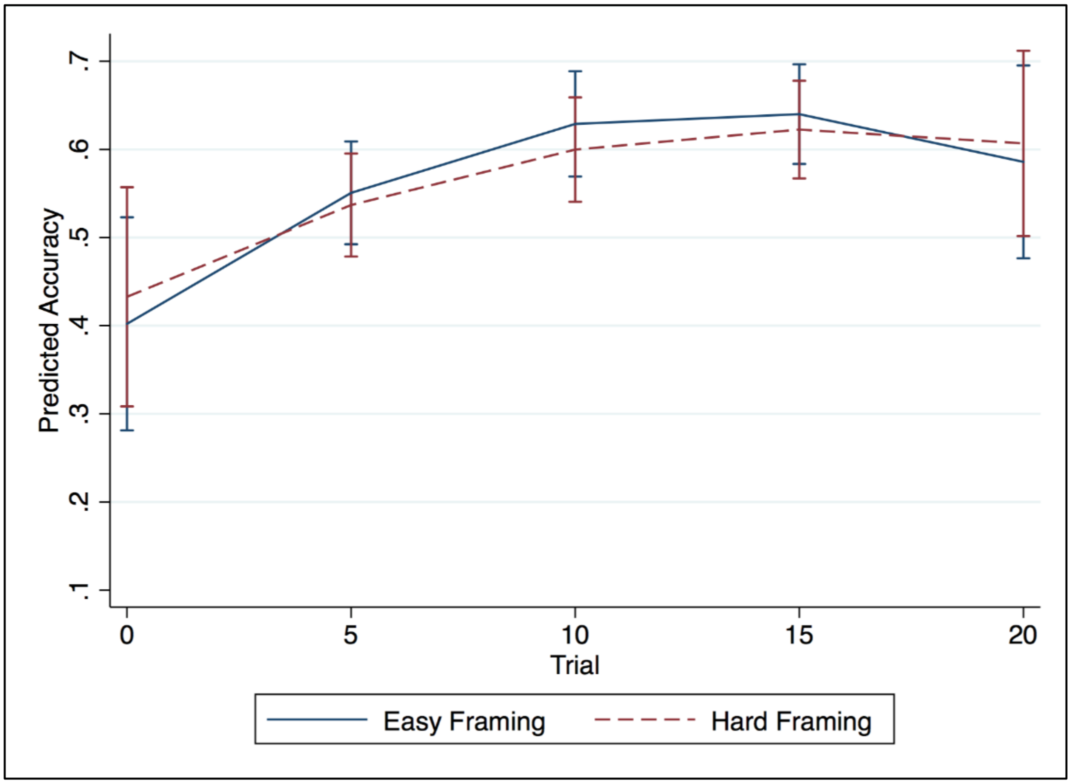

Figure 5: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback)

We then focused on accuracy and confidence. As shown in Figure 5, accuracy had similar curvilinear relationships with experience in both conditions. Participants became more accurate quite quickly in the first trials. However, their rate of learning slowed in later trials and approached zero after trial 15.

As shown in Figure 6, the relationship between confidence and experience tended to be positive in both conditions. However, it was curvilinear for participants told the task would be easy, and linear for participants told the task would be hard.

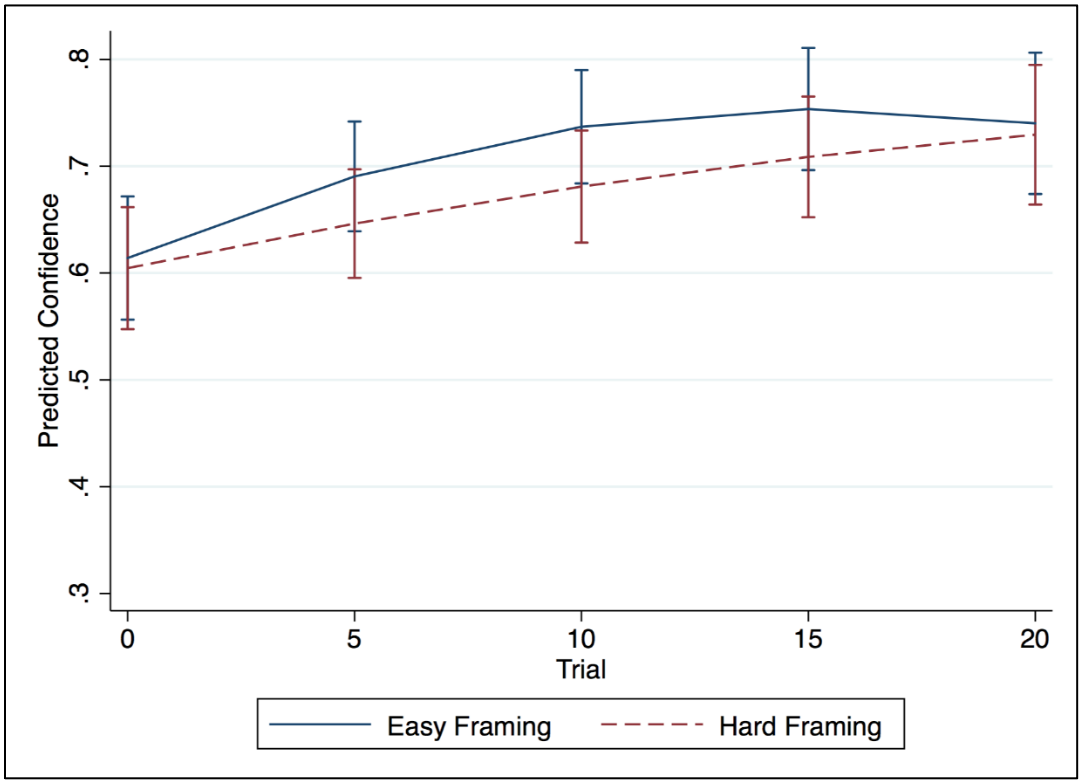

Figure 6: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 2 (With Feedback)

Finally, we tested Hypothesis 1, which stated that accuracy would have a lagged effect on confidence in a subsequent trial. Acknowledging that confidence and accuracy could have reciprocal effects with one another, we tested this using a cross-lagged effect analysis based on the maximum likelihood estimator (Allison, Williams, Moral-Benito, 2017; Moral-Benito, 2013). This method allows us to control for unobserved, time-invariant confounders (e.g., individual differences). It also allows us to estimate models with endogenous regressors, which helps make inferences about causal direction (see Allison, 2009; Arellano, 2003). Our results suggest that confidence at time $t+1$ was affected by both the lagged confidence from time t (b = .41, $p < .001$) and accuracy from time t (b = .16, $p < .001$). Similarly, overconfidence at time $t+1$ was affected by both the lagged overconfidence from time t (b = .33, $p < .001$) and accuracy from time t (b = .02,

$p < .001$). In contrast, confidence at time *t* did not affect accuracy at time *t+1* ($b = .002$, $p = .824$).

In sum, considering how accuracy and confidence changed over time in both groups, the evolution of overconfidence in this study was substantially different from that in Study 1. Feedback tended to reduce overconfidence in all participants, which can be explained mostly by their accuracy: The growth in accuracy exceeded the growth in confidence over time. It is interesting to note that feedback stopped reducing overconfidence (i.e., overconfidence approached zero) at trial 15, where there was an inflection point. Finally, we found that giving a correct answer, as well as receiving feedback about it, tended to increase confidence and overconfidence in later trials.

3.3 STUDY 3

3.3.1 Method

The goal of Study 3 was to test whether the results of Study 2 could be replicated using a different sample.

Study 3 was held at the Center for Decision Research at the University of Chicago's Booth School of Business and was carried out in accordance with the recommendations of the University's Social and Behavioral Institutional Review Board. As with previous studies, we included the framing difficulty manipulation in the introduction and then asked participants to take part in the same 20-item estimation task.

In addition to the different sample, another difference between Studies 2 and 3 was that—due to operational reasons—we did not offer incentives for calibration in Study 3. This also allowed us to have a simpler introduction message. We included the same outcome feedback in Study 3 as we did in Study 2.

Ninety-three subjects participated in the experiment. Their average age was 29 ($SD = 13$). Only 20% of participants answered the gender question. Participants were offered US$1 for every 5 minutes they spent on the study. The experiment was computer-based and held in two laboratories at the Center for Decision Research. As was the case in Studies 1 and 2, participants were randomly assigned into two different framing difficulty groups (easy or hard). The measures we used were the same as those included in Study 2: experience, level of confidence, actual performance (accuracy), and overconfidence.

3.3.2 Results and Discussion

When overconfidence was used as the dependent variable (Table 3, left panel), experience had a significant quadratic effect on overconfidence. Feedback had a stronger effect in initial trials than it did in the final trials. Also, as was the case in Study 2, Figure 7 shows that there was an inflection point around trial 15. Thus, the answer to Research Question 3, which asked whether experience under feedback had a quadratic relationship with overconfidence, is 'yes'.

As was the case in Study 2, there was no significant quadratic Experience^2 x Difficulty Framing interaction. This finding was similar when using accuracy as the dependent variable (binary logistic link function; Table 3, center panel). We did find a main effect of the framing manipulation on confidence (Table 3, right panel). However, this difference in confidence did not translate into a significant difference in overconfidence. Thus, as was the case in Study 2, feedback did not moderate the effect of difficulty framing on overconfidence (Research Question 4). Regarding Hypothesis 1, we did find a lagged effect of accuracy on confidence (b = .01, $p < .05$) but not on overconfidence (b = .01, p = .11). Additionally, results revealed no lagged effect of confidence on accuracy.

We should point out two further interesting results from this study, both of which are related to confidence. First, contrary to Study 2, confidence did not increase with feedback over time. Second, as can be seen by comparing Figures 6 and 9, confidence (and thus overconfidence) was substantially larger in the Chilean sample compared to the Chicago sample. In fact, from these figures we learn that, while participants in the

Chilean sample (Study 2) were overconfident by the end of the experiment, participants in the Chicago sample (Study 3) were not.

	Overconfidence	Accuracy	Confidence
Hard Framing	-0.061	-0.14	-0.097*
Experience (Trial)	-0.034***	0.163*	0.000
Hard Framing x Experience	-0.002	0.02	0.007
Experience x Experience	0.001***	-0.006*	0.000
Hard Framing x Experience x Experience	0.000	-0.00	-0.0003
Constant	0.162***	-0.582	0.543***
N	1860	186	1860
AIC	-2694.0	2504.	-2677.7
BIC	-2639.1	2549.	-2627.9
ICC	0.835	0.030	0.814

Note. For framing, hard = 1, easy = 0.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3

Study 3: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors. For confidence and overconfidence, interaction between Experience and

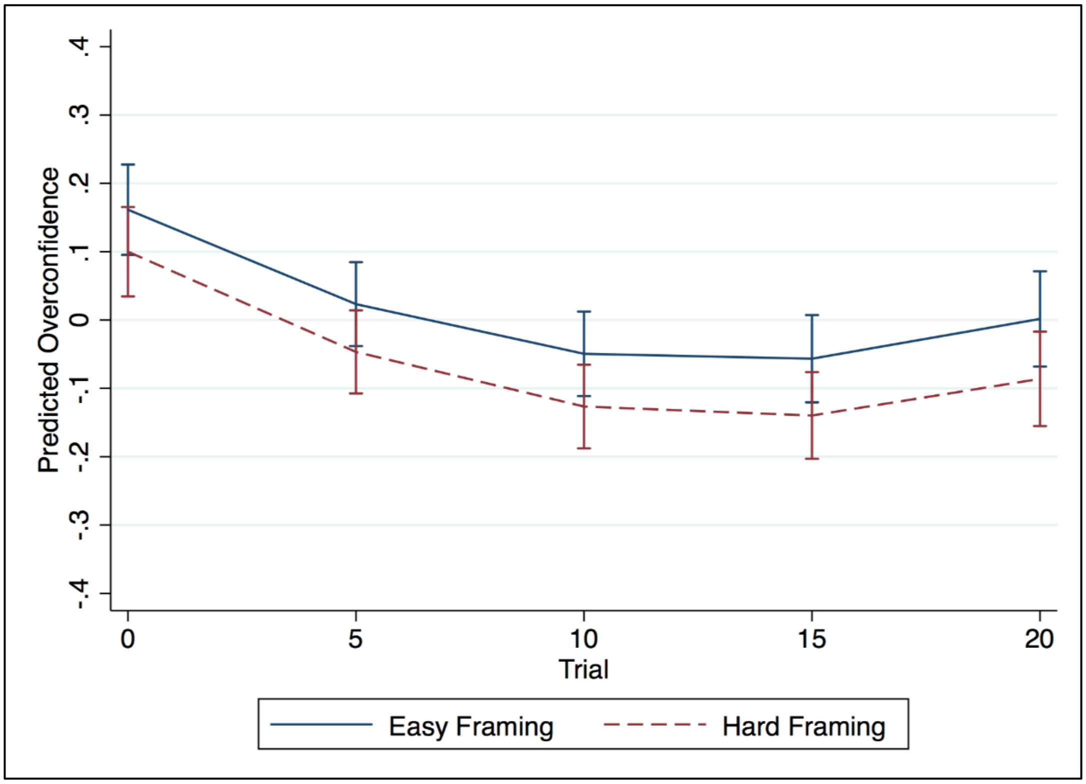

Figure 7: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (No Feedback)

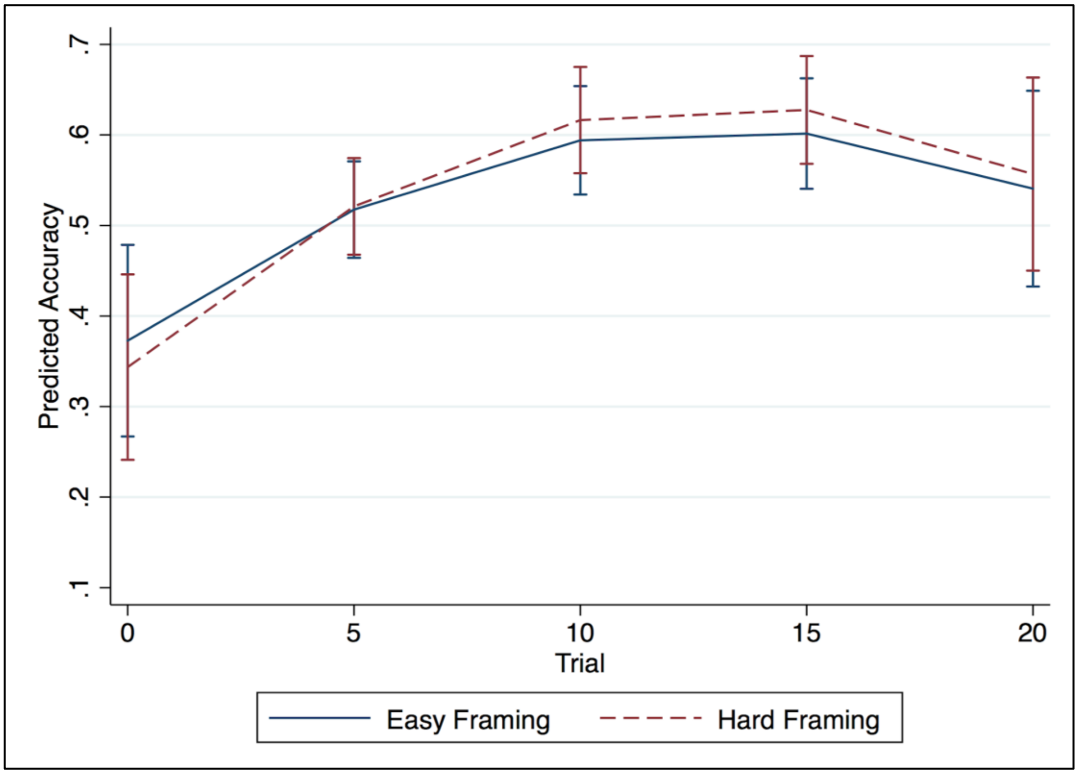

Figure 8: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With feedback).

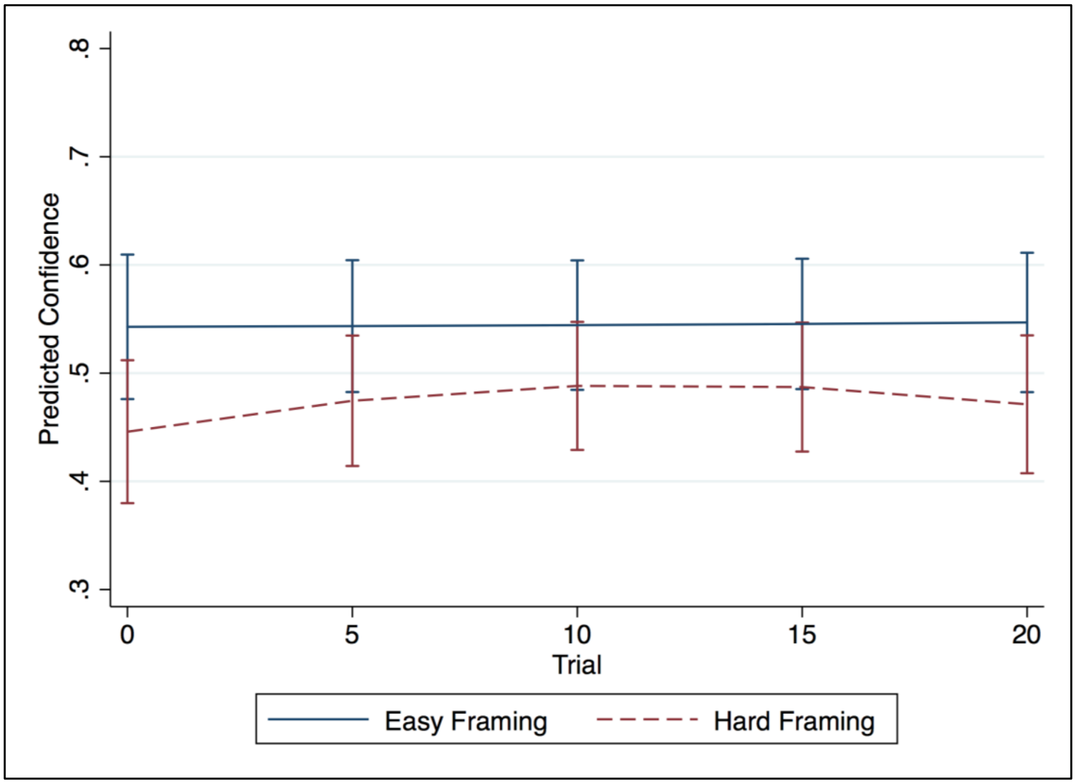

Figure 9: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With Feedback)

In sum, we found that experience with feedback reduced overconfidence in a non-linear way in Study 3. As was the case in Study 2, we found that after trial 15 there was no further benefit from providing feedback. However, there were two important differences between these results and those of Study 2. Participants in Study 3 (Chicago sample) were substantially less confident than those in Study 2 (Chilean sample). Also, while confidence tended to increase over time in the Chilean sample, confidence remained constant across trials in the Chicago sample. Given these differences, we conducted a fourth study. We sought to determine whether these disparities in findings were due to differences in the sample or differences in the procedure.

3.4 STUDY 4

3.4.1 Method

3.4.1.1 Participants and Procedure

Seventy-two participants, who were working professionals completing a management degree at the Pontificia Universidad Católica de Chile a large Chilean university, took part in this experiment. The mean age of participants was 36 (*SD* = 7.9) and 28% were female. As was the case in Studies 1 and 2, the experiment was conducted online and each participant received the link through their personal email. Participants also were randomly assigned to the two difficulty framing manipulations (hard vs. easy). As noted, in Study 4 we used the same experimental manipulation as Study 3. We also used the same measures for all variable as Studies 2 and 3.

3.4.2 Results and Discussion

We first focused on overconfidence. Table 4 shows that results are, in terms of significance, closer to results in Study 2 than those in Study 3.

	Overconfidence	Accuracy	Confidence
Hard Framing	-0.185***	0.54	-0.073
Experience (Trial)	-0.066***	0.271**	-0.004
Hard Framing x Experience	0.040***	-0.13	0.012***
Experience x Experience	0.003***	-0.011**	0.000
Hard Framing x Experience x Experience	-0.002***	0.00	-0.000**
Constant	0.345***	-1.015**	0.617***
N	1440	144	1440
AIC	-3059.5	1927.	-2441.3
BIC	-3006.9	1969.	-2393.9
ICC	0.763	0.01	0.747

Note. For framing, hard = 1, easy = 0.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4

Study 4: Multilevel regression with overconfidence, accuracy, and confidence as outcomes, and experience, experience squared, framing, and their interactions as predictors.

Figure 10 shows a strong effect of the framing difficulty manipulation on overconfidence in early trials. Feedback reduced the difference in overconfidence across conditions. This effect was strong at the beginning but gradually decreased. Stated differently, there was a strong curvilinear trend in the relationship between feedback and overconfidence for participants told the task would be easy. This effect was lower for those told the task would be hard. Thus, the answers to Research Questions 3 and 4, which ask whether feedback had a curvilinear effect on overconfidence and whether feedback moderates the effect of framing on overconfidence over time, are both ‘yes’.

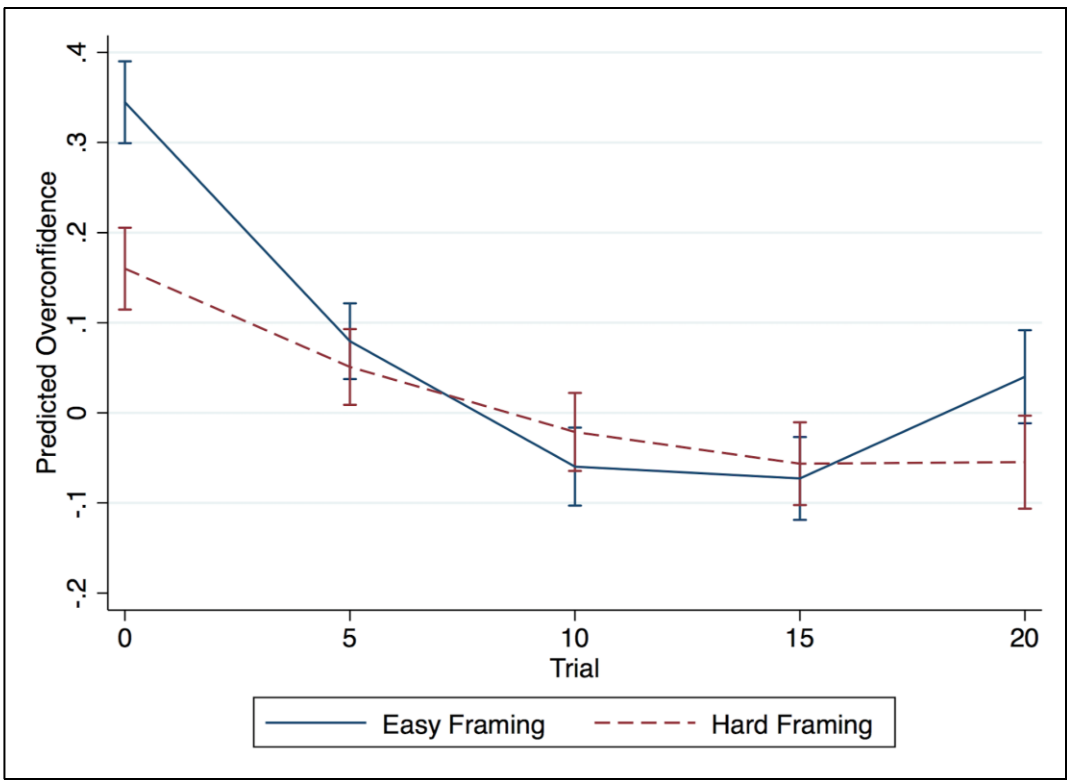

Figure 10: Overconfidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 4 (With Feedback)

Regarding accuracy, as can be seen in Table 4 and Figure 11, feedback reduced the difference in overconfidence between participants in both difficulty framing conditions. Learning existed in both groups, but was more curvilinear for participants in the hard framing condition than for those in the easy framing condition. As for confidence, the initial differences tended to quickly disappear after trial 5. Finally, as was the case in Study 2, we found that accuracy had a lagged effect on confidence (*b* = .01, $p < .05$) and overconfidence (*b* = .01, $p < .01$).

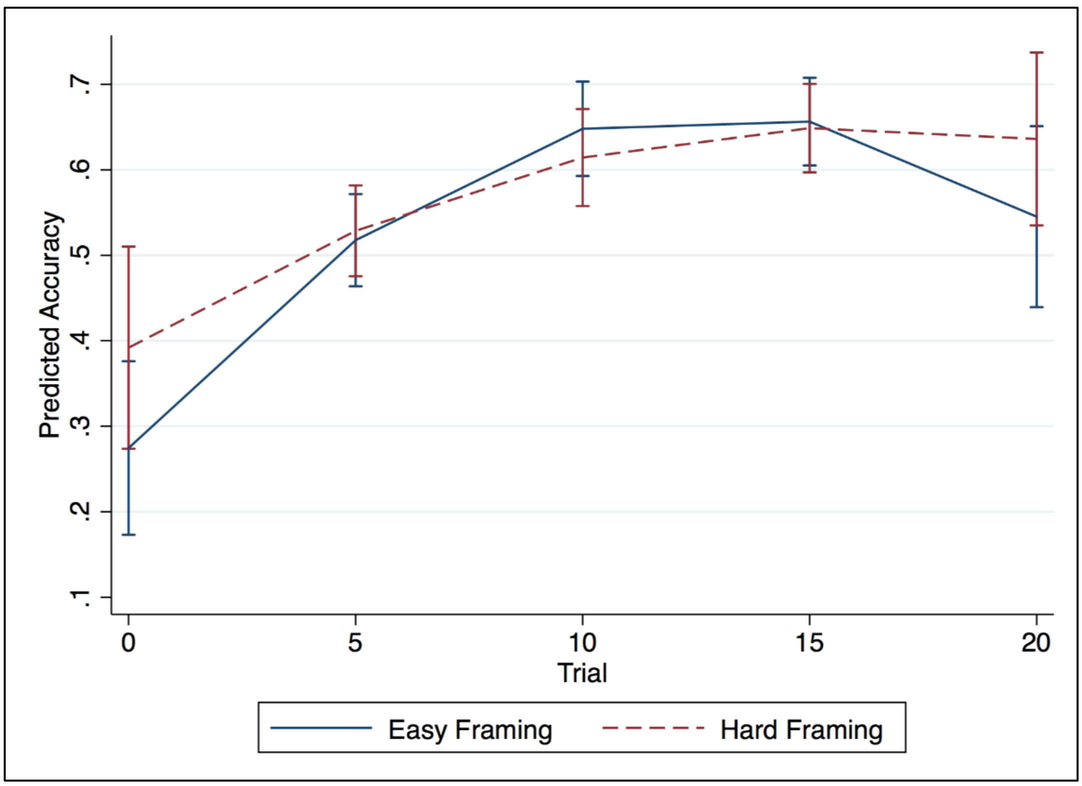

Figure 11: Accuracy as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 4 (With feedback).

In sum, the results of Study 4 were more similar to those of Study 3 than those of Study 2. Experience with feedback decreased overconfidence in a quadratic pattern, but more so for participants told the task would be hard. That is, participants in the easy framing condition showed initially more overconfidence than those in the hard framing condition, but feedback quickly leveled these differences. As was the case in Studies 2 and 3, feedback did not reduce overconfidence beyond trial 15 for participants in either condition. Finally, note that the confidence slope (but not its intercept) was more similar to that in Study 3 than that in Study 2. We offer an explanation for this in the General Discussion.

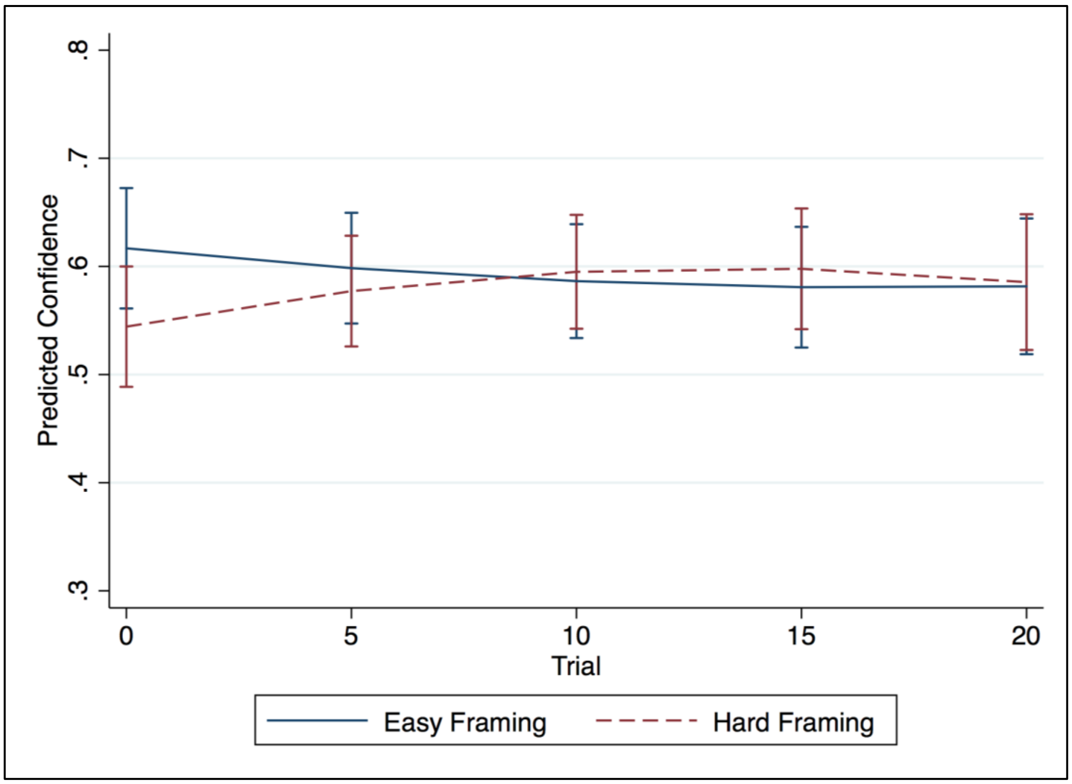

Figure 12: Confidence as a Function of Experience through 20 trials for people told the task would be easy vs. hard, Study 3 (With Feedback)

4. GENERAL DISCUSSION

Our main purpose in this study was to understand how overconfidence evolves over time. In order to do this, we sought to answer more specific research questions. First, we examined whether having experience with a task decreased overconfidence. Second, we studied whether experience reduced the effect of framing the difficulty of a task on overconfidence. Third, we examined whether feedback decreased overconfidence over time. More specifically, we studied whether the effect of feedback on overconfidence was incremental (small but constant across trials) or exponential (an initial strong effect that declines across trials). Fourth, we tested whether accuracy at time t had lagged effects on confidence at time $t +1$. To test these questions, we conducted four laboratory experiments across two different samples using a relatively simple word-counting task.

4.1 General description of the main findings

Our results suggest that experience alone does not clearly reduce overconfidence (Study 1). In our experiment, the effects of experience depended on the beliefs participants held, which in turn were based on how the task was framed. Participants who were told the task would be easy slightly decreased their large initial overconfidence over time. Those told the task would be hard reduced their overconfidence in the initial trials; however, their overconfidence increased after 10 trials, perhaps due to a fatigue (or discouragement) effect. Accuracy, rather than confidence, tended to explain these results.

Feedback clearly decreased overconfidence over time (Studies 2 - 4). This effect was strong in the initial trials and faded in subsequent trials. Indeed, by trial 15, feedback no longer reduced overconfidence. This could have two explanations. First, feedback may have been informative only at the outset of the experiment. Stated differently, feedback gave an initial anchor of the number of words contained in each text passage; subsequently, this information may have been redundant. (Note, however, that accuracy levels by the end of the task were far from perfect). Second, this could be the result of a floor effect. For example, participants in Studies 3 and 4 (Figures 7 and 10) were already presenting almost perfect calibration (no overconfidence) by trial 10. As such, there may not have been much room for improvement.

Finally, cross-lagged analyses generally showed that, among those who received feedback, accuracy at time t had an effect on confidence and overconfidence at $t+1$. Participants who had a correct estimate were more likely to increase their confidence and overconfidence in a subsequent trial. Confidence and overconfidence had no effect on accuracy.

4.2 Discrepancies across studies

We also found some results that were not consistent across experiments (Table 5 shows the ways in which studies differed). The first discrepancy in the studies' results was related to the effect of the framing difficulty manipulation on overconfidence.. Participants in Studies 2, 3, and 4 all received feedback. However, whereas the framing difficulty manipulation did not have an effect on overconfidence in Studies 2 and 3, it did have an effect in Study 4. In the initial trials of Study 4, participants in the easy framing condition showed greater overconfidence than those in the hard framing condition. These differences quickly disappeared with feedback over time. We suspect this discrepancy is related to the different samples (Chicago vs. Chilean) as well as differences in the experiment procedure (incentives for calibration; see below).

Table 5: Ways in which studies differed: Feedback, Location and Reward.

Study	Feedback	Location	Reward
1	No	Chile	Yes
2	Yes	Chile	Yes
3	Yes	USA	No

4	Yes	Chile	No

A second difference across studies relates to confidence levels. Participants in the Chicago sample in Study 3 had substantially lower confidence levels across trials (*M* = .51; *SE* = 0.01) than the Chilean samples in Study 2 (*M* = .69; *SE* = 0.01) and Study 4 (*M* = .59; *SE* = 0.01). Also, confidence levels tended to increase over time in Studies 1 and 2, especially in the easy framing condition. This increase was not present in Studies 3 and 4.

We believe there are two reasons for these differences in confidence levels. First, we told participants in Studies 1 and 2 that they would be rewarded based on their Brier score. In other words, we were giving an incentive for showing good calibration. In Studies 3 and 4, we did not include this reward. As a result, in Studies 1 and 2, participants may have stated riskier (bolder) confidence levels, especially by the end of the experiment. Confidence levels were less important in Studies 3 and 4. Indeed, perfect calibration (zero overconfidence) was never achieved in Study 1 (no feedback) or Study 2 (with feedback), unlike in Studies 3 and 4.

Second, we showed the same text passages to both the Chicago and Chilean samples across experiments, meaning that the text passages that the Chicago sample assessed were in Spanish. Prior to beginning the experiment, we believed that this would have no effect on the estimates; after all, participants had to complete a word-counting task, not a comprehension task. However, Spanish words may have induced additional disfluency (Alter, Oppenheimer, Epley, & Eyre, 2007) among Chicago participants, thus reducing their confidence.

4.3 Implications

Our study has two implications for the overconfidence literature. First, our study is consistent with previous research suggesting that experience, unless coupled with unequivocal feedback or calibration training, does not necessarily reduce overconfidence (Arkes, 2001; Kahneman & Klein, 2009; Mandel & Barnes, 2014; Shanteau, 1992; Sanchez & Dunning, 2018; Tetlock, 2017). This is particularly interesting considering the literature on decisions from experience (Wulff, Mergenthaler-Canseco, & Hertwig, 2018). According to much of this literature, there is a gap between how individuals make decisions from descriptions vs. experience. One assumption is that decisions from experience tend to be less biased than those made from descriptions. For example, whereas context effects are quite strong when individuals are asked to make decisions from a described set of options (e.g., Connolly, Reb, & Kausel, 2013), these effects are greatly reduced when individuals have experience with the options' attributes (Hadar, Danziger, & Hertwig, in press). Our results show, though, that mere experience is not enough to de-bias an individual's initial overconfidence when they engage in a task. Experience coupled with feedback (or training), however, does de-bias an individual's initial overconfidence (Mellers et al., 2015). Likewise, our study suggests that feedback over time tends to make pre-task beliefs disappear, but this is less clear in tasks where experience is not coupled with feedback.

A second implication is the importance of considering both confidence and accuracy levels when studying overconfidence (i.e., overestimation; Moore & Healy, 2008). In other words, when overconfidence is reduced, is it because of an increase in accuracy, a decrease in confidence, or both? This identification problem is pervasive in any measure utilizing difference scores, such as overestimation (i.e., confidence minus accuracy; Edwards, 1994; Gigerenzer, 1991). In our studies, feedback reduced overconfidence mainly due to increased accuracy.

We believe this is an important issue. A number of researchers assume that low overconfidence is an indicator of meta-knowledge: The degree to which we are aware about what we know and what we do not know (Kahneman, 2011; Russo & Schoemaker, 1992). This may be the case in some situations, but our results—especially those from Studies 3 and 4—may be explained by a simple reason. One explanation is that participants were unsure of how they were doing throughout the experiment. Indeed, their confidence trend in these studies was essentially constant across trials (Figures 9 and 12). Another, more plausible explanation, is that participants in these studies did not care about providing reliable estimates; their confidence levels may have been mostly random noise.[2] In any case, both explanations suggest that participants may not have gained any new insights regarding their skill or affected their meta-knowledge. Their low overconfidence simply stems from the fact that they provided more correct answers. In our view, changes in overconfidence levels may be attributed to several reasons, and (lack of) meta-knowledge is only one of them.

4.4 Limitations

One issue that may limit the generalizability of our results is related to the task. As noted above, the task was quite simple. This may explain the quadratic relationship between overconfidence and trials under feedback. In other words, participants likely learned to reduce their overconfidence (or achieved calibration) only in the early trials because of the simplicity of the task. Relatedly, this may explain why the task description had little to no effect on overconfidence, especially in Studies 2 and 3. A recent paper suggested that experience in easy tasks, compared to descriptions, had

---

[2] To further explore this issue, we examined an indicator of the quality of confidence estimates: the relationship between confidence and accuracy across trials within-subjects (this is akin to what have been termed *slope* [Yates, 1982] or *monitoring resolution* and *discrimination accuracy* [Liberman & Tversky, 1993; Koriat, 2012]). We did this by conducting pooled feasible generalized least squares regressions (PFGLS; Cameron & Trivedi, 2010) using the within (i.e., 'fixed-effects') estimator. We used confidence as the dependent variable and accuracy as the independent variable. In Study 2, we found that the within-subject relationship between confidence and accuracy was significant ($b = .028$, $p < .001$). In contrast, in Studies 3 and 4, it was not ($b_{Study\ 3} = .006$, $p = .260$; $b_{Study\ 4} = .008$, $p = .124$).
We conclude that the quality of the confidence estimates was better in Study 2 (when good calibration was rewarded) than in Studies 3 and 4 (when good calibration was not rewarded).

a stronger influence on judgments than experience in hard tasks (Weiss-Cohen, Konstantinidis, Speekenbrink, & Harvey, 2018). As such, in more complex tasks, the evolution of confidence may have a different shape.

A second limitation has to do with how we rewarded calibration. On the one hand, when we did not reward participants for calibration (Studies 3 and 4), their confidence levels were of lower quality, as our analyses reported in footnote 2 suggest. On other hand, when we did reward participants (Studies 1 and 2), they also reported bolder confidence levels. This may have to do with the structure of the incentive we offered. Recall that we offered the equivalent to a US$75 cash prize only for the top performers of each study based on their Brier scores. Given that there was only a single cash prize, participants may have stated higher confidence levels in these studies. In retrospect, a more sensitive strategy would have been giving incentives on a trial-to-trial basis; however, we believe that the structure of incentives is a very important issue that future research should address when studying confidence and overconfidence.

4.5 Conclusion

Notwithstanding these limitations, our set of studies shows the importance of examining how confidence and overconfidence evolve over time. We found that experience does have an effect on—but does not necessarily reduce—overconfidence over time. Feedback, however, quickly reduces overconfidence primarily due to improved accuracy. We hope that our research can inspire more studies examining these issues.

REFERENCES

Allison, P. D. (2009). *Fixed effects regression models* (Vol. 160). SAGE publications.

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. *Socius*, *3*, 2378023117710578.

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.

Arellano, M. (2003). *Panel data econometrics*. Oxford university press.

Arenius, P., & Minniti, M. (2005). Perceptual variables and nascent entrepreneurship. *Small business economics*, *24*(3), 233-247.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In *Principles of forecasting* (pp. 495-515). Springer, Boston, MA.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer Science & Business Media.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, *116*(1), 261-292.

Berger, J., & Pope, D. (2011). Can losing lead to winning?. *Management Science*, *57*(5), 817-827.

Bernardo, A., & Welch, I. (1997). *A Theory of Overconfidence and Entrepreneurship*. working paper, UCLA Anderson School.

Bollen, N. P. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, *42*(3), 683-708.

Brier, G. W. (1950). The statistical theory of turbulence and the problem of diffusion in the atmosphere. *Journal of Meteorology*, *7*(4), 283-290.

Budescu, D. V., & Johnson, T. R. (2011). A model-based approach for the analysis of the calibration of probability judgments. *Judgment and Decision making*, *6*(8), 857.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata press.

Connolly, T., Reb, J., & Kausel, E. E. (2013). Regret salience and accountability in the decoy effect. *Judgment and Decision making*, *8*(2), 136.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121-136.

Deaves, R., Lüders, E., & Schröder, M. (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization*, *75*(3), 402-412.

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management*, *20*(3), 683-689.

Feng, L., & Seasholes, M. S. (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets?. *Review of Finance*, *9*(3), 305-351.

Forbes, D. P. (2005). Are some entrepreneurs more overconfident than others?. *Journal of business venturing*, *20*(5), 623-640.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European review of social psychology*, *2*(1), 83-115.

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1-36.

Gloede, O., Menkhoff, L., & Waibel, H. (2015). Shocks, individual risk attitude, and vulnerability to poverty among rural households in Thailand and Vietnam. *World Development*, *71*, 54-78.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual review of sociology*, *26*(1), 441-462.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, *15*(8), 534-539.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in cognitive sciences*, *13*(12), 517-523.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American psychologist*, *64*(6), 515.

Kahneman, D., Lovallo, D., & Sibony, O. (2011). Before you make that big decision. *Harvard business review*, *89*(6), 50-60.

Kausel, E., Perez-Cotapo, V., Zhang, S, & Reyes, T. (2018). Can overconfidence be induced? Overestimation and perceived task difficulty. *Manuscript submitted for publication.*

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27-44.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes*, *79*(3), 216-247.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*(2), 162.

Mandel, D. R., & Barnes, A. (2014). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences*, *111*(30), 10984-10989.

Mandel, D. R., & Barnes, A. (2018). Geopolitical Forecasting Skill in Strategic Intelligence. *Journal of Behavioral Decision Making*, *31*(1), 127-137.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., ... & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of experimental psychology: applied*, *21*(1), 1.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., ... & Ungar, L. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, *10*(3), 267-281.

Menkhoff, L., Schmeling, M., & Schmidt, U. (2013). Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization*, *86*, 92-101.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & cognition*, *37*(7), 1026-1039.

Mishra, K. C., & Metilda, M. J. (2015). A study on the impact of investment experience, gender, and level of education on overconfidence and self-attribution bias. *IIMB Management Review*, *27*(4), 228-239.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, *115*(2), 502.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics*, *31*(4), 451-472.

Pompian, M. M. (2006). Behavioral finance and wealth management. *How to build optimalportfoliosforprivate clients*, 28.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences*, *23*(1), 125-133.

Rogers, T., & Moore, D. (2014). The Motivating Power of Under-Confidence:'The Race is Close But We're Losing'.

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan management review*, *33*(2), 7.

Schulz, J. F., & Thöni, C. (2016). Overconfidence and career choice. *PloS one*, *11*(1), e0145126.

Shanteau, J. (1992). How much information does an expert use? Is it relevant?. *Acta Psychologica*, *81*(1), 75-86.

Slaughter, J. E., & Kausel, E. E. (2013). Employee selection decisions. *Judgment and Decision Making at Work*, *57*.

Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational behavior and human decision processes*, *83*(2), 282-309.

Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, *119*(2), 295-300.

Tetlock, P. E. (2017). *Expert political judgment: How good is it? How can we know?*. Princeton University Press.

Waldman, D. A., & Yammarino, F. J. (1999). CEO charismatic leadership: Levels-of-management and levels-of-analysis effects. *Academy of management review*, *24*(2), 266-285.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological bulletin*, *144*(2), 140.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, *30*(1), 132-156.