OVERCONFIDENCE AMONG BEGINNERS

A Dissertation

Presented to the Faculty of the Graduate School of Cornell University

In Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy

by

Carmen Julia Sanchez

May 2020

OVERCONFIDENCE AMONG BEGINNERS

Carmen Julia Sanchez

Cornell University, 2020

Abstract: There is a vast literature that examines differences in overconfidence between people, depending on their knowledge level. In this line of research, the tasks used do not involve learning. As people approach these types of tasks, they generally have some knowledge about their abilities. Some are better, while others are worse, and perceptions of their abilities are assessed. Although there is a lot of research exploring differences in overconfidence depending on knowledge level, there is relatively limited research looking at how overconfidence changes in people as they learn. Much of my work is devoted to exploring the exact shape and timeline of that overconfidence, with an emphasis on probabilistic learning, in which people learn to read cues from the environment to predict some outcome.

To explore this line of work first, I mapped out the trajectory of performance, confidence, and overconfidence as people engage in novel probabilistic tasks (Chapter 2). I found that people generally start with rather modest self-assessments, with their confidence tracking performance rather well, but then a problem develops. With just a little learning, confidence sky-rockets far above accuracy, a phase I refer to as the "beginner's bubble" of overconfidence.

There are several reasons for the development of this beginner's bubble. First, overconfidence develops because people formulate faulty theories about how to approach a task, and once people have an idea about how to perform a task, even if it is wrong, it produces overconfidence (Chapter 2). Second, when people first engage in a task, they pay close attention to feedback to guide their decisions and are appropriately calibrated about their abilities.

However, at a certain point, the experienced learner no longer carefully pays attention to feedback to discover why they made mistakes. In fact, overconfidence is not driven by feedback. It is best explained by how quickly people make decisions. The mere act of making a series of decisions, when similar ones have been made before, drives overconfidence (Chapter 3).

In exploring mechanisms that lead to the development of the beginner's bubble, I have also identified individual differences in overconfidence resulting from a specific behavior, the propensity to "jump-to-conclusions" (Chapter 4). Those who jump-to-conclusions are defined as those who collect less information to reach conclusions in problem solving tasks.

People who jump-to-conclusions have divergent trajectories of overconfidence as they learn. They have an elevated confidence curve as they are learning, lower levels of overall performance, and are thus more overconfident. These individuals are also more likely to endorse false beliefs, like conspiracy theories. In this line of research, I have also modified metacognitive interventions that are ordinarily designed to reduce delusional beliefs, to quell overconfidence in a probabilistic learning task, without negatively impacting learning.

Biographical Sketch

Carmen Sanchez graduated from the University of Texas at El Paso with a bachelor's degree in business administration, with a concentration in management. Prior to attending Cornell University, she worked for ten years in banking, most of this time was spent as a national bank examiner at the Office of the Comptroller of the Currency, an agency of the U.S. Department of Treasury. One day she decided to quit her perfectly good job to pursue a more fun career in academia. After graduating from Cornell, she will be starting her academic career as an Assistant Professor of Organizational Behavior at the University of Illinois at Urbana-Champaign.

Acknowledgements

I would first like to express my deepest gratitude to my mentors, my adviser Tom Gilovich and co-adviser David Dunning. I thank Tom for his optimism, unwavering encouragement, and guidance throughout all these years. In addition, I would like to thank David for his never-ending realism, believing in me, especially when I didn't believe in myself, and for being the most hilarious person I know. Thank you both for your kindness, infinite patience, and the research opportunities and freedom that have allowed me to reach this point.

Besides my advisers, I would like to thank my committee members - Melissa Ferguson and Alex Ophir - for their contributions to my professional development and guidance throughout all these years. I am also extremely grateful to Robert Ricketts, and Jason and Shannon Rinaldo for encouraging me to pursue a career in research, and for listening to all my whining.

Thanks to my family for the many sacrifices they made coming to this country, and their constant reminder of them. I thank my father for always pushing me to work harder and, my mother who values education more than anyone I know, even though she was never allowed the opportunity of one.

I am also grateful to have the support of my very wonderful friends Angela and Franccesca. I would also like to say thank you to my friends and collaborators, Xi, Fangming, and Jonas, that make doing research a lot more fun. Thanks to my dogs, Bob and Peter Parker, for all your cuteness and judgment free listening and understanding. Lastly, I would like to thank Daniel who loves his job more than anyone that I know and wouldn't let me settle for anything less.

Portions of the research presented in Chapter 4 were funded by the Templeton Foundation through the Thrive Center program on the Science of Intellectual Humility awarded to David Dunning. Furthermore, funding for this dissertation was also provided for by the Cornell Provost Diversity Fellowship.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 1: Introduction

As former baseball pitcher Vernon Law once put it, experience is a hard teacher because it gives the test first, and only then provides the lesson.

Perhaps this observation can explain the results of a survey sponsored by the Association of American Colleges & Universities. Among college students, 64% said they were well prepared to work in a team, 66% thought they had adequate critical thinking skills, and 65% said they were proficient in written communication. However, among employers who had recently hired college students, less than 40% agreed with any of those statements. The students thought they were much further along in the learning curve toward workplace success than their future employers did (Hart Research Associates, 2015).

My research focuses on how overconfidence changes as people learn. To be a beginner is to be susceptible to undue optimism and confidence. Much of my work is devoted to exploring the exact shape and timeline of that overconfidence, with an emphasis on probabilistic learning, in which people learn to read cues from the environment to predict some outcome. For example, people must rely on multiple signals from the environment to predict which company's stock price will rise the most, which applicant will do the best job, or which illness a patient is suffering from. These can be hard tasks—and even the most expert of experts is at times going to make the wrong prediction—but a decision is often required in many settings.

Outline of Chapters 2 - 4

Chapter 2

One factor that plays a role in judgment quality is the time course of decisions. We all begin as beginners. How good are we, as beginners, at tracking our progress on a task we barely know anything about?

In this chapter, I present a reproduction of a paper published in 2018, along with coauthor David Dunning in the *Journal of Personality and Social Psychology*, titled "Overconfidence among beginners: Is a little learning a dangerous thing?" Across six studies, I mapped out the trajectory of performance, confidence, and overconfidence. In four initial studies, participants completed multicue probabilistic learning tasks (e.g., learning to diagnose "zombie diseases" from physical symptoms). In 2 additional studies I tested whether I could find real-world echoes of the pattern I found in the initial four studies across the life course. I find a consistent and replicable pattern I refer to as the beginner's bubble of overconfidence throughout all my studies.

Chapter 3

There are several reasons for the beginner's bubble and in this chapter, I discuss another mechanism that relates to the development of overconfidence, fluency. When people first engage in a task, they pay close attention to feedback to guide their decisions and are appropriately calibrated about their abilities. However, at a certain point, the experienced learner no longer carefully pays attention to feedback to discover why they made mistakes. In fact, overconfidence is not driven by feedback. It is best explained by how quickly people spend making decisions. The mere act of making a series of decisions, when similar ones have been made before, drives overconfidence.

In this chapter, I present a reproduction of a paper that is in press, along with coauthor David Dunning in *Decision*, titled "Decision fluency and overconfidence among beginners."

Chapter 4

Lastly, in exploring mechanisms that lead to the development of the beginner's bubble, I have also identified individual differences in overconfidence resulting from a specific behavior, the propensity to "jump-to-conclusions." Those who jump-to-conclusions are defined as those who collect less information to reach conclusions in problem solving tasks.

Interestingly, jumping-to-conclusions has been extensively studied for decades, but mostly on schizophrenia patients. In schizophrenia research, patients who "jump-to-conclusions" in probabilistic reasoning tasks tend to display impaired decision-making and delusional belief. I examined whether jumping-to-conclusions (JTC) was similarly associated with decision impairments in a nonclinical population, such as reasoning errors, false belief, overconfidence, and diminished learning. In this line of work, I have also developed an intervention that dampens overconfidence as people engage in a novel probabilistic learning task.

In this chapter, I present a reproduction of a paper that is a revise and resubmit, along with coauthor David Dunning at the *Journal of Personality and Social Psychology*, titled "Jumping-to-conclusions: Implications for reasoning errors, false belief, knowledge corruption, and impeded learning."

Reference

Hart Research Associates. (2015). Falling short? College learning and career success. *Association of American Colleges and Universities.* Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Chapter 2: Overconfidence Among Beginners: Is a Little Learning a Dangerous Thing?[1]

A little learning is a dangerous thing;

Drink deep, or taste not the Pierian spring;

There shallow draughts intoxicate the brain,

And drinking largely sobers us again.

--Alexander Pope (1711)

Of all the errors and biases people make in self and social judgment, overconfidence arguably shows the widest range in its implications and the most trouble in its potential costs. Overconfidence occurs when one overestimates the chance that one's judgments are accurate or that one's decisions are correct (Dunning, Heath, & Suls, 2004; Dunning, Griffin, Milojkovic, & Ross, 1990; Fischhoff, Slovic, & Lichtenstein, 1977; Moore & Healy, 2008; Russo & Schoemaker, 1992; Vallone, Griffin, Lin, & Ross, 1990).

Research shows that the costs associated with overconfident judgments are broad and substantive. Overconfidence leads to an overabundance of risk-taking (Hayward, Shepherd, & Griffin, 2006). It prompts stock market traders to trade too often, typically to their detriment (Barber & Odean, 2000), and people to invest in decisions leading to too little profit (Camerer & Lovallo, 1999; Hayward & Hambrick, 1997). In medicine, it contributes to diagnostic error (Berner & Graber, 2008). In negotiation, it leads people to unwise intransigence and conflict

(Thompson & Loewenstein, 1992). In extreme cases, it can smooth the tragic road to war (Johnson, 2004).

To be sure, overconfidence does have its advantages. Confident people, even overconfident ones, are esteemed by their peers (Anderson, Brion, Moore, & Kennedy, 2012). It may also allow people to escape the stress associated with pessimistic thought (Armor & Taylor, 1998), although it does suppress the delight associated with success (McGraw, Mellers, & Ritov, 2004). However, as Nobel laureate Daniel Kahneman has put it, if he had a magic wand to eliminate just one judgmental bias from the world, overconfidence would be the one he would banish (Kahneman, 2011).

In this manuscript, we study a circumstance most likely to produce overconfidence, namely, being a beginner at some task or skill. We trace how well confidence tracks actual performance from the point where people begin their involvement with a task to better describe when confidence adheres to performance and when it veers into unrealistic and overly-positive appraisal—that is, how closely the subjective learning curve fits the objective one.

Popular culture suggests that beginners are pervasively plagued by overconfidence, and even predicts the specific time-course and psychology underlying that overconfidence. According to the popular "four stages of competence" model, widely discussed on the Internet (e.g., Adams, 2017; Pateros, 2017; Wikipedia, 2017), beginners show a great deal of error and overconfidence that dissipates as they acquire a complex skill. At first, people are naïve about their deficits and are best described as "unconscious incompetents," not having adequate awareness of just how unskilled they are. In the academic literature, this would be described as the Dunning-Kruger effect (Dunning, 2011; Kruger & Dunning, 1999), a situation in which people are so unskilled they lack the very expertise necessary to recognize their shortcomings.

However, with more experience, people pass into a "conscious incompetence" phase, in which they perform poorly but recognize it. Upon further practice, people graduate to the "conscious competence" phase in which they are aware of how to complete a task successfully, but still needs a good deal of deliberative thought to succeed. Finally, people reach "unconscious competence," in which a skill becomes second nature, requiring little to no conscious thought.

The Beginner's Bubble Hypothesis

In the research contained herein, although we agree that beginner status and overconfidence are often related, our reading of the psychological literature leads us to propose a different pattern of development from that described by the four stage model.

As a main hypothesis, we propose instead a pattern that looks like a "beginner's bubble." Specifically, we suggest that people begin their career at some task by being quite cautious and unconfident in their decisions, but that they quickly become overconfident—the beginner's bubble—before going through a "correction" phase in which confidence flattens while performance continues to improve. In essence, we flip the order of the unconscious and conscious incompetence phases noted above and suggest that people do not begin in a Dunning-Kruger state, but acquire it after a little experience. As expressed in the famous Alexander Pope quotation that begins this manuscript, when it comes to overconfidence, a little learning is a dangerous thing, leading to overinflated self-perceptions of expertise after a few shallow draughts of experience that begins to deflate slowly only with continued consumption of experience and learning.

Theoretical Rationale

We propose this specific pattern of confidence and overconfidence, first, because it better matches both our intuition and the literature about how overconfidence would develop among beginners in a complex task. Rank beginners, we assert, will show very little overconfidence, if indeed any confidence in their skill. Imagine that we assigned our readers to start tomorrow to authenticate works of art for the Louvre, to judge which applicants are the best bets to repay their bank loans, or sign up as a homicide detective. We doubt anyone with zero experience at any of these tasks would claim much confidence as they start. People would likely have no theory or strategy about how to approach the task. Consistent with this assertion, extant studies on perceptions of skill learning (Billeter, Kalra, & Loewenstein, 2011) and memory performance (Koriat, 1993) suggest that rank beginners often underrate or appropriately rate their future performance at a task.

However, after some experience with the task, even a little bit, people will rapidly grow confident and even overconfident about their judgments. This will particularly be true in multi-cue probabilistic learning tasks, in which people must mull over cues from the environment to make predictions about uncertain events, such as deciding which company's stock will rise the most, which job applicant will do the best job, or which illness their patient is suffering from. Cues can be helpful in reaching the right decision, but not with complete certainty.

This is a type of task that characterizes many of complex challenges people face in life (Brunswick, 1943; Estes, 1976; Little & Lewandowsky, 2012). However, although there is voluminous data on probabilistic learning, to our knowledge there is a slim amount of work comparing objective learning curves (performance) with subjective ones (confidence) (e.g., Fischer & Budescu, 2005; Sieck & Yates, 2001), and none focusing specifically at confidence as participants approach a task as an absolute beginner. Usually, instead, there is a study or practice period before researchers begin assessing confidence (Fischhoff & Slovic, 1980).

We assert that beginners will quickly develop overconfidence in probabilistic learning tasks because they are exuberant theorizers and pattern matchers. They will take feedback and outside information to quickly make inferences and spin beliefs about how to make correct decisions (Sieck & Yates, 2001). Much work in psychology has shown for decades that people are very comfortable taking impoverished data, and such small portions of it, to reach confident theories about events and how they should react (Dunning, 2012; Risen, Gilovich, & Dunning, 2007; Heider & Simmel, 1944). They can read meaningful patterns in putatively random or meaningless data (Chapman & Chapman, 1969; Guthrie, Weber, & Kimmerly, 1993; Rabin, 2002; Ono, 1987), or even recruit information from past life experience in the absence of data (Fischhoff & Slovic, 1980).

The problem with this exuberant theorizing is that small portions of data usually contain a substantial degree of noise and potentially misleading information. The know-how beginners generate exuberantly may be more apparent than real. As such, confidence based on that theorizing will race ahead, but accurate judgment will be much slower to the race. To be sure, as people continue in on the task, the mistaken portions of their theorizing will be pointed out to them. They will make errors that they learn from. As such, their performance will improve, but it will generate no more overconfidence as they revise and prune their theories away from mistaken notions toward more accurate ones.

Research on the "belief in small numbers" supports this analysis, showing how people are insensitive about how much data they have before reaching their conclusions, assuming that very small samples of data are good indicators of what the world is really like when in fact those early pieces of data may contain a good deal of noise (Benjamin, Rabin, & Raymond, 2015; Griffin & Tversky, 1992; Tversky & Kahneman, 1971; Williams, Lombrozo, & Rehder, 2013). Often, the first piece of information people see has an undue weight on subsequent theorizing (Asch, 1946;

Jones, Rock, Shaver, Goethals, & Ward, 1968; Kelley, 1950), and can prevent them from recognizing true patterns evident in the world (Kamin, 1968; Yarritu, Matute, & Luque, 2015). In short, people quickly build theories based on the "strength" of the evidence they see early on in a task, failing to temper their theorizing given the small "weight" they should give to the evidence because of how little there is of it (Griffin & Tversky, 1992).

Supportive Empirical Evidence

Importantly, if one looks at empirical work on skill and error among beginners, one sees a pattern suggestive of our account of overconfidence. Beginners often appear to start learning a new skill cautiously and with few errors. They are risk-averse and vigilant. It takes a little while for confidence to build, as evidenced by the time-course of errors they typically show. The most widely-known example of this is the so-called "killing zone" in aviation (Craig, 2013; Knecht, 2013). Beginning pilots are appropriately cautious in the cockpit, not crashing their planes at any great rate. However, as they accumulate more flight hours, they become more dangerous, experiencing fatal crashes at increasing rates until roughly 800 flight hours, after which crash rates begin to decline slowly. In short, flight errors often attributed to overconfidence or carelessness follow more of a beginner's bubble pattern that develops over time than one associated with the four stages model, which would suggest the most overconfident errors would be among absolute beginners to aviation.

Medical errors follow the same pattern: Initial wariness gives way to a bubble of overconfidence and careless error, which then declines. Some spinal surgeries involve guiding a robotic device to place stabilizing screws into spinal vertebrae. The first five surgeries a beginner completes require supervision, after which beginners are on their own. However, surgeons do not spike in errors immediately after their supervision is over. Instead, their greatest spike in misplacement of robotic screws does not typically occur until between their 16th and 20th surgeries (Schatlo et al., 2015). Furthermore, physicians with a medium amount of training have higher rates of false negative diagnoses than both experts and beginners when performing gastrointestinal endoscopies (O'Callaghan, Miyamoto, Takahashi, & Fujita, 1990).

On the other end of the organism, dentists with a mere interest in a type of specialized dentistry exhibit higher error rates than those with both no knowledge and those with high levels of expertise (Avon & Wood, 2010). In addition, medical students are more underconfident in their diagnoses in clinically challenging cases than are more senior medical residents or doctors with at least two years of experience after medical school, even though diagnostic accuracy rises reliably with seniority. Medical students are overconfident in only 25% of cases where their diagnoses "misalign" with the correct diagnosis, whereas residents and practicing physicians show the same tendency on 41% and 36% of cases, respectively (Friedman et al., 2005).

Beyond Beginners

Beyond a beginner's bubble, we remain agnostic about where the relationship between confidence and accuracy will end up, when learning finally gives way to expertise. In general, the higher the knowledge level the more closely confidence matches performance. Not surprisingly, some research finds that experts tend to outperform novices across many domains and are also better calibrated in their confidence estimates (Ericcson, 1991; Wallsten and Budescu, 1983). However, other research finds that even highly trained professionals remain overconfident (Hazard & Peterson, 1973; Cambridge & Shreckengost, 1978; Oskamp, 1962; Von Holstein, 1972; Moore, 1977; Neale & Bazerman, 1990; Wagenaar & Keren, 1986, Hynes & Vanmarcke, 1976; McKenzie, Liersch, & Yaniv, 2008). In addition, it seems that access to a larger and richer knowledge base either makes people better calibrated or, makes decisions easier to justify, inducing overconfidence (Gill, Swann, & Silvera, 1998; Oskamp, 1965; Swann & Gill, 1997). As such, although we make strong predictions about the advent of confidence among beginners, we refrain from making equally strong predictions about where people will end up as they acquire additional expertise.

Overview of Studies

In all, we examined the beginner's bubble hypothesis across six studies. In each, we examined how confidence versus competence developed as people gained more experience at a complex task.

Our primary focus in the first four studies was on probabilistic learning. In two initial studies, we examined whether beginner confidence and overconfidence arose in the specific pattern we predicted as people gained experience, and incrementally became more accurate, in two different probabilistic learning tasks. In the third study, we added incentives to further ensure that the confidence estimates participants provided represented their true beliefs.

In the fourth study, we examined whether exuberant theorizing underlay the pattern of confidence we observed. We asked people in a mock medical diagnosis task to describe the principles or strategies they followed as they diagnosed their "patients." We predicted that people would quickly develop self-assured theories that inspired confidence but which contained a good deal of error. Further experience, however, would prune some of that error away while confidence steadied or deflated. As such, we predicted that the pattern of confidence we observed would be explained by the time-course of the theories that people developed as they gained experience.

Finally, in Study 5a and 5b, we switched to a real-world task of some complexity, examining extant data on financial literacy across the lifespan to see if it followed the same pattern of subjective and objective learning curves we found in the laboratory. We expected self-confidence in financial literacy to rise markedly among young adults, but then flatten until later in the life course. Real financial literacy, however, would show a slower and more incremental rise across age groups.

Study 1: The Development of Overconfidence

In Study 1, our aim was to understand how people assess their judgments when learning to make decisions whose outcomes are predictable but uncertain. Participants completed a novel medical diagnostic task, similar to one used in previous research (McKenzie, 1998). Participants were asked to imagine they were medical residents in a post-apocalyptic world that has been overrun by zombies. Over 60 repeated trials, they diagnosed possible zombie infections from information on eight different symptoms that could indicate unhealthy patients, receiving feedback about their accuracy after each trial. Similar to the real world, all symptoms attached to ill health had varying probabilities; diagnosis was, thus, based on fallible clues.

We predicted that participants would incrementally learn how to diagnose patients more accurately, thus showing a predominantly linear learning curve. Confidence in those judgments, however, would follow a path that is consistent with our beginner’s bubble hypothesis. Participants would start well-calibrated or even underconfident in their diagnoses, but would quickly develop confidence levels that outstripped their levels of accuracy. That confidence level, however, would soon flatten. In short, whereas accuracy would rise in linear fashion, confidence would follow a non-linear path. In regression terms, it would follow at least a negative quadratic trend, with a quick rise that then deflated.

Method

Participants. Forty participants were recruited from Amazon’s Mechanical Turk crowdsourcing facility. Participants received $3 for their participation. In addition, they had the chance to win an additional $3 if they achieved an overall accuracy level of 80% in the medical diagnosis task. The sample consisted of 60% men and 40% women.

To enhance statistical power, we exploited within-subject designs, focusing primarily on how confidence and accuracy unfolded for each participant through time. Given this circumstance, we used a rather crude estimation procedure to compute our needed sample size due to uncertainties we faced in the sizes of our predicted effects and complexities of calculating power in the specific data analysis strategy we adopted (Hayes, 2006). We anticipated that our effects, all within-subject, would be moderate in size ($d = .5$), given pilot data, and so calculated the sample size needed to capture such an effect in a within-subject comparison. At a sample size of 31, we calculated an 80% chance of capturing a significant finding ($\alpha = .05$), but rounded up our initial sample size to 40 participants to be conservative. In subsequent studies, we raised our target sample sizes to 50 to raise power to near 95%.

Procedure. Participants were instructed that they would be taking part in a hypothetical medical diagnosis scenario. Two strains of zombie disease had broken out across the world, TS-19 and Mad Zombie Disorder (MZD). Luckily, a team of virologists had developed medication that cured affected patients, but only if accurately diagnosed. Failing to use the appropriate medication could be potentially fatal.

Participants were instructed that they had been rescued by the National Guard and provided refuge at the Centers for Disease Control and Prevention, where they had become a medical resident under supervision of renowned Dr. John Walker. They were being trained in zombie disease detection and treatment. As part of their training they were about to see patients. They were further instructed that all of these patients had either TS-19, MZD, or neither. TS-19 and MZD could not occur at the same time in a patient. Both diseases had common symptoms but there are varying probabilities of the symptoms associated with the two illnesses. Some symptoms were distractions, not associated with either illness. Participants were then given a short quiz to ensure they understood the task they were about to perform. They were provided immediate feedback about the accuracy of their choices on the quiz.

After the quiz, participants were told that Dr. Walker needed to leave town for a couple of days to train other residents. Participants would have to diagnose the next 60 patients on their own. They would receive feedback after each diagnosis about their accuracy. They were reminded that there was a 25% chance of any symptom being present yet the patient not being sick. Also, there is a chance that the patients were sick even when not exhibiting symptoms.

Participants were then presented sixty patient profiles, one at a time. Each profile listed eight symptoms and stated whether each symptom was present or absent in the current patient. Participants diagnosed each patient as having TS-19, MZD, or neither. They also reported how confident they were of their decision would prove accurate. Specifically, they were instructed:

> Please report how confident you are in this decision. What's the chance that you are right, from 33% to 100%? Mark 33% if you think it's just as likely that you are wrong as you are right (i.e., it's 33-33-33 that I'm right). Mark 100% if you are absolutely sure that you are right; there's no chance that you are wrong. Mark 66% if you think the chance that you are right is 2 out of 3. Mark whichever probability best indicates the specific chance that you are right.

After participants reported their confidence for each case, they were given immediate feedback on their performance. Feedback included the right diagnosis and repeated the symptom profile presented for that patient. Participants were allowed to keep written records of the information they received and the decisions they made. In fact, participants were instructed that it might be helpful to create a table with all of the symptoms and illnesses and to place a checkmark next to the symptoms as they are going through the patients. A sample empty table was provided to them with all symptoms listed in a vertical fashion on the left side of the table, and the possible diagnoses (TS-19, MZD, and neither) were listed on the top of the table in a vertical manner.

Materials. Patient profiles listed eight physical symptoms (congestion, itching, brain inflammation, abscess, swollen glands, rash, fever, and glossy eyes) that were potentially indicative of a zombie disease. Two of the eight were diagnostic of TS-19 disease (e.g., congestion was present in 80% of such patients, but only 20% present in MZD or 25% of healthy patients). Two of the eight were diagnostic of MZD (e.g., glossy eyes were present in 80% of such patients, but only 25% of TS-19 sufferers and 25% of healthy patients). One symptom was equally associated with both syndromes (i.e., abscess was present in 70% of both syndromes, but only 25% of healthy patients), and three symptoms were nondiagnostic (e.g., swollen glands were present in 20% of patients suffering either syndrome and 25% of those who were healthy).

To create the patient profiles, symptoms were randomly assigned to the patient profiles via pre-arranged probabilities. Participants were not aware of these probabilities while they were performing the task. They simply knew that the probabilities of these symptoms occurring varied by diagnoses, not all patients would present with the same symptoms and highly diagnostic symptoms would not always be present. Specific patient profiles were presented in four different sequences to counterbalance individual cases with the order in which they were confronted.

Results and Discussion

Data from 2 participants were excluded because they never moved their confidence rating for any individual case from the default of 33%. It was presumed they skipped this measure.

Accuracy. To assess whether participants learned, we conducted a logistic mixed model analysis (random-intercept, random-slope) assigning experience (i.e., trial number) as a fixed variable and participant as a random variable.[2] We then examined whether experience predicted participant accuracy. Consistent with our hypothesis, participants increased in accuracy across the 60 diagnoses they made, b = .0054, se_b = .0025, p = .032, OR = 1.01. As Figure 1 (left panel) shows, participants started roughly 54% accurate and ended around 64% accurate. As a cautionary analysis, we then added a quadratic experience term in a second analysis to see if there was a significant non-linear effect of experience on learning. The quadratic term was not significant, z = -0.02, *ns*.

Confidence. Overall, participants proved overconfident in their diagnoses. To compare confidence and accuracy, we recoded diagnoses in which participants were accurate as 100% and those they were wrong as 0%. We then submitted diagnoses to a mixed model analysis in which type of response, confidence or accuracy, was coded as 1 or 2, respectively, nested within participants in a random intercept, random slope model. Confidence overall (M = 69.3%) far exceeded accuracy (M = 60.0%), $t(37.0)$ = 3.70, p = .005, η_p^2 = .27.

But how did that overconfidence develop with experience? We next examined whether confidence mirrored the linear trend in learning or departed from it. We predicted that confidence would follow a curvilinear path, and so subjected confidence ratings to a mixed model regression analysis including both a linear and quadratic term for experience as fixed effects and participants (random intercept, random slopes) as a random variable. Both terms were

[2] Preliminary analyses assigning specific case profile (i.e., the 60 particular cases that participants diagnose) as a random variable produced either models that did not converge or ones that produced results virtually identical to those reported in the text. Thus, we did not take include case profile in our analyses.

significant (see Table 1), and the overall model was a better fit (as measured by *BIC*) than a simple linear model.

As an exploratory analysis, we also repeated the analysis, this time including a cubic term for experience (along with nesting the cubic trend within participants via random-slopes). This more complicated model returned an unexpected but significant cubic trend (see Table 1), with this model demonstrating a slightly better fit than our initial one. In sum, and as Figure 1 (left panel) shows, it appears that as people learn, they do not start confident but there is a rapid increase in confidence that eventually levels off, as we predicted. However, and unpredicted, confidence then begins to increase again as people gain extensive experience with a task.[3]

Overconfidence. We finally focused on patterns of overconfidence, more for descriptive purposes than for inferential ones. For both confidence and accuracy for each diagnosis trial, we calculated the fitted value for that trial and its standard error. Then, for each trial, after converting the data for accuracy from binary to continuous format, we then subtracted the fitted accuracy values in the linear model from the fitted confidence levels in the cubic model described above. Thus, for each of the 60 trials, from the first to the last, we had an estimate of the degree of overconfidence expressed. We then calculated a standard error for that overconfidence estimate as:

$$SE_{OC} = \text{square root } (SE_C^2 + SE_A^2 \ - 2 \text{ x } r_{AC} \text{ x } SE_C \text{ X } SE_A)$$

In the equation, OC = overconfidence, C = confidence, A = accuracy and r_{AC} is the correlation between accuracy and confidence. Using that standard error, we calculated a 95% confidence interval for the degree of overconfidence that participants displayed.

Again, we did this analysis more to describe the pattern of overconfidence participants

[3] We also explored quartic models in the first four studies, just to be sure, and found that only one produced a significant quartic term and also an improved *BIC*. In other studies, *BIC* actually regressed, indicating a poorer fit.

displayed while they gained experience rather than conduct any inferential tests. With that caution in place, as seen in Figure 1 (right panel), participants appear not to be clearly overconfident on average until the tenth case they diagnosed, and their overconfidence increased until their 27^{th} case, where overconfidence sat at nearly 13%. However, after that case, their confidence, as predicted, sagged down to roughly 10% by trial 49, after which it rose unexpectedly back again to 12% by the end of 60 patients.

Taken together, these findings provide initial evidence for a curvilinear relationship in overconfidence as people learn. People initially start with low levels of confidence that rapidly spike to an inappropriate "beginner's bubble" level, which then levels off for a while their learning continues, only to restart a rise again later.

Study 2: Conceptual Replication

Study 2 was similar to the study above in that participants learned a novel task whose outcome varied with uncertainty. However, we changed the materials and sought to replicate the previous results, including the unanticipated cubic trend in confidence. In this study, participants were asked to imagine they were researchers that had just invented two lie detection devices, based on similar technologies. Both machines were sensitive to different types of criteria that were known to be associated with lying. They needed to choose which machine would best detect the lying given the criteria they had.

Method

Participants. Fifty participants were recruited from Amazon's Mechanical Turk. They received $3 for their participation.

Materials. Similar to Study 1, four different orders of case profiles were created to counterbalance for order effects.

Procedure. Participants were instructed that they would be taking part in a hypothetical research and product development scenario. They had just invented two types of lie detector devices, the Doodad and the Thingymabob. To understand how they work, the two devices need to be tested. Both of the machines are sensitive to different types of criteria that have been known to be associated with lying (e.g., sweating). That is, they detect lying but they did so in different ways.

Before they could make millions selling their machines, participants needed to determine which criteria helped both of the lie detection devices best detect when people were lying. Certain criteria mean the Doodad would be better at detecting lying and others meant the Thingymabob would be better. Some of the criteria are useful for both machines. Other criteria, however, did not detect lying very well and thus were useless to both machines. The participant's task was to find the combinations of criteria that proved useful for each machine. Participants were then given a short true or false quiz to ensure they understood the task they were about to perform.

After the quiz, participants were instructed they were ready to start testing their lie detection devices. They had strapped both lie detection devices to sixty individuals who had been instructed to lie. Most of these individuals would feel rather uncomfortable lying. As such, they would exhibit different behavioral signs of lying, such as heavy breathing or sweating. Participants needed to figure out which machine would detect deception the best given the criteria that each test individual exhibited. Each test case profile listed the eight criteria and stated whether each was present or absent in the case. Participants stated whether the Doodad, Thingymabob, or neither would best detect the lying. They also reported how confident they were their decision would prove accurate, using same measure and instructions for confidence used in this study as in Study 1. They were given immediate feedback on their performance. Feedback designated the correct device, and repeated the profile presented for that liar. As in Study 1, participants were instructed that it might be helpful to create a table with all of the devices and to place a checkmark next to the criteria as they are going through the liars. They were also provided with a sample table.

Results and Discussion

One of our case profiles was faulty and was excluded in the analysis from this study. Therefore, our task comprised 59 repeated trials. We conducted identical analyses as in study 1 for accuracy and confidence and replicated our results. Accuracy rose with experience in a linear fashion (see Figure 2, left panel), b = .010, se_b = .0023, p < .001, *OR* = .01. The quadratic component of accuracy was not significant, b = -.003, p = .079. Confidence overall significantly exceeded accuracy (*M*s = 64.2 and 52.7, respectively, $t(49.0) = 4.62$, p < .001, η_p^2 = .31. It also followed the same curvilinear cubic relationship observed over trials in Study 1(see Table 1 and Figure 2), thus replicating the results of Study 1, b = .0004, p < .001, for the cubic term.

Study 3: Incentivizing Confidence Estimates

In Study 3, we provided incentives not only for accuracy but also for valid measures of confidence. In doing so, we adopted a procedure, the Becker-DeGroot-Marschak method (Becker, DeGroot, & Marschack, 1964) to induce participants to provide confidence estimates that best characterized their true beliefs about whether or not their diagnoses were accurate. In economic terms, this procedure aimed at ensuring that confidence estimates were “incentive compatible,” that is, they were designed to motivate participants to tell the truth about their confidence while removing any pressures to be strategic other than telling the truth (Schotter & Trevino, 2015).

Method

Participants. Fifty undergraduate students from Cornell University participated for course credit. They had the chance to win up to $3 if they achieved certain accuracy levels across all trials. Additionally, they had a chance to win an additional $5 depending on how they reported their accuracy.

Procedure. The materials and procedures used in this study were taken from Study 1, except for a few alterations. Notably, after participants completed the quiz that ensured they understood the probabilistic learning task, they learned they could win an additional $5 either in a lottery or if one of their diagnoses was accurate. They were told they would see sixty patient profiles. For each, they would report their confidence in their decision from 33% (I'm guessing) to 100% (I'm sure) based on the instructions given below.

Specifically, participants were told that at the end of the zombie diagnosis task one of their diagnoses would be selected randomly to see if they would win the additional $5. The confidence level they expressed for that diagnosis, however, would determine whether they would win the $5 based on the accuracy of their diagnosis or instead in a random lottery that they could switch to. The key to the lottery was that we would not announce the chance of winning until it was time to play. The question the participant had to decide for themselves was, would they rather bet that their diagnosis was right or instead on the lottery for each possible chance of winning we might name (e.g., 40%, 50%, 60%). In other words, for each diagnosis, they were asked to indicate the probability level at which they would rather switch from betting on their diagnosis to taking their chances on the lottery.

For example, if they were 70% confident in their diagnosis, that meant that they wanted to bet on their diagnosis instead of any lottery with a chance of winning of 70% or less, but that they would want to switch to the lottery if it offered a chance of winning that was 71% or above. Similarly, a 40% confidence meant they wanted to make the switch from betting on their diagnosis to the lottery if the chance of winning at the lottery were 41% or higher. Participants were instructed that to increase the likelihood of winning they should be as honest as possible in how they reported their confidence. Participants then answered several questions to ensure they understood how to report their confidence to earn the most money. Feedback on the accuracy of each question on the quiz was provided immediately.

Past work has shown that this form of probability assessment prompts people to provide confidence estimates that better represent their true beliefs rather than beliefs contaminated by other strategies or biases, such as risk aversion (Blavatskyy, 2009; Trautmann & van de Kuilen, 2015). We constrained the "bet" to win additional money to only one randomly selected diagnosis to prevent "portfolio management" (e.g., hedging on some bets and then mixing in some risky ones).

Participants then engaged in the same 60-case probabilistic learning task that was used in Study 1. We then randomly selected the same diagnostic case for everyone and played the additional bet, paying off those participants who won.

Results and Discussion

We used the same analyses as those used in the previous two studies and replicated our findings. Consistent with the previous studies, there was a significant linear trend in accuracy across the 60 trials, b = .009, se_b = .002, p < .001, OR = 1.01. Participants started at 55% accuracy and ended at roughly 68% (see Figure 3, left panel). No quadratic trend emerged when tested, z = -1.16, *ns*.

Overall, confidence exceeded accuracy significantly, as tested via the mixed model analysis used in Study 1, *M*s = 70.2 and 61.4, respectively, $t(49.0) = 4.27$, $p < .001$, $\eta_p^2 = .27$. A cubic model (see Table 1) produced the best fit and also yielded a significant cubic trend, $b = .0003$, $se_b = .0001$, p = .004, $\eta_p^2 = .15$ (see Figure 3, left panel). In sum, Study 3 replicated the results of the previous studies with a careful, incentive-compatible measure of confidence.

Study 4: Theoretical Exuberance as an Underlying Mechanism

Study 4 was designed to test our proposed mechanism for the beginner's bubble, that people actively construct theories of prediction too exuberantly, forming quick but self-assured ideas of how to approach our probabilistic learning task based only on small shards of data. Those early pieces of data contain a substantial degree of noise, and so any insights based on them contain a good deal of spurious content, serving more as apparent knowledge than authentic know-how. After an initial exuberance phase, continued experience chips away at misleading components of those theories while reinforcing their accurate pieces, leaving people as incrementally more accurate as they gain more experience, but with flat confidence as their theories are revised. At some point, those incremental revisions do give way to more definitive theories, leading to the tail-end rise in confidence.

In short, we predicted that exuberant theorizing underlies people's confidence in their judgments, and importantly follows a cubic trend over experience explaining the pattern of confidence we observed in the first three studies. The accurate component of those theories, however, is more linear and incremental, leading to the simpler objective learning curve in accuracy that we observed.

In a replication of the zombie diagnosis task, we tested our exuberant theorizing account by asking participants to report the theories underlying their diagnoses before they began the task and then after every twelve trials. More specifically, we assessed whether participants had formed a theory about the outcome each symptom was connected to (versus stated they did not know) as well as how confident they were in that inference. We then aggregated these reports into an overall index of theory development. Importantly, our methods allowed us to separate accurate theorizing (i.e., the participant correctly connected the symptom to the right diagnosis) from erroneous theorizing (i.e., the participant attached the symptom to the wrong outcome). We predicted that accurate theorizing would display more of an incremental linear trend, and thus explain the linear trend seen in the previous three studies concerning diagnostic accuracy.

Method

Participants.

Forty-nine participants were recruited from Amazon's Mechanical Turk crowdsourcing facility.

Procedure.

In this study we used the same zombie task from Study 1, save one major addition. To gauge the degree of theory development, we added a task to test how quickly participants developed partial to full-blown theories about how medical symptoms were connected to possible diagnoses. To do this, we embedded questions at six points throughout the study. At these time points, participants answered 16 questions regarding their medical theories about diagnosing zombie disease. Participants were presented with the eight individual symptoms used in the task, and asked for each whether it indicated a MZD diagnosis, a TS-19 one, either (i.e., the person was ill, but the symptom did not distinguish which specific illness was present), neither (i.e., the person is healthy), or was irrelevant. If they indicated an answer, they then rated their confidence in that answer from 1(not at all) to 5(certain) that they were right. Finally, participants were also allowed to answer for each symptom that they did not know.

From these responses, we constructed a scale of theory development. If participants gave an answer, we gave them a score based on their confidence (i.e., a score from 1 to 5). If participants stated they did not know, they received a score for that symptom of 0. We then summed all participant scores across all eight symptoms. As such, a person's theory development score could range from 0 (refused to provide any theory about any symptom) to 40 (offered conclusions for all eight symptoms of which they were completely certain). This overall theory development score could be bifurcated into two components. One part of the score represented theory development for those symptoms in which participants gave a correct answer about the outcome the symptom indicated. The other was for those instances in which the participant gave an erroneous answer. Participants reported their theories first just before beginning the diagnosis task, and then again after their 12th, 24th,36th, and 48th trials, with the last report occurring right after the 60th and final trial.

Results and Discussion

Two participants never varied their diagnostic confidence estimates from the default setting of 33%, suggesting they were ignoring the measure. Their data were omitted.

Confidence and Accuracy

We replicated the impact of experience on accuracy and confidence in the diagnosis task (see Figure 4). Accuracy again rose in an incremental linear fashion, $b = .006$, $se_b = .002$, $p = .010$, $OR = 1.01$, with no further curvilinear trend detected when added to the model, $z = -0.47$, *ns*. Confidence again was best explained by a model including linear, quadratic, and cubic trends, $b = .0003$, $se_b = .00009$, $p < .001$, $\eta_p^2 = .20$, for the cubic trend (see Table 1).

Unlike other studies, overall confidence ($M = 61.6$) did not significantly exceed accuracy ($M = 58.3$), $t(40.0) = 1.21$, *ns*, $\eta_p^2 = .02$. However, the time course of confidence as participants gained experience mirrored that of the previous studies. Specifically, participants did not start out as overconfident (see Figure 4, right panel), but as 6.5% underconfident. Confidence then rose much more quickly than did accuracy over early cases with confidence exceeding accuracy by roughly 5% by case 28. Overconfidence then flattened down slightly to 4% by case 44 as accuracy continued to rise, and then began to rise again to roughly 8% by case 60.

Theory Development

Would underlying theory development follow the same cubic time course as confidence, with an early burst leading to a flat retrenchment, then a final rise? A look at Figure 5, which tracks theory development over experience, suggests that it did. To confirm, we subjected overall theory development scores to a mixed-model random-intercept, random-slope analysis in which order was entered as a fixed variable, with participant as a random variable. We decomposed order effect into its linear (weighting order as -5, -3, -1, 1, 3, 5), quadratic (weighting = 5, -1, -4, -4, -1, 5) and cubic (-5, 7, 4, -4, -7, 5) trends. All trends were significant, F = 26.34, 13.85, 40.07, η_p^2 = .36, .23, .47, respectively, ps < .001. Although all three trends were significant, it is interesting to note that the biggest trend we found in this analysis was the cubic one.

We provide Figure 6 as another way to depict the time course of theory development as participants experienced the medical diagnosis task. The figure depicts changes in theory development between theory probes, rather than the degree of theory development at a particular theory probe. The figure clearly shows that participants developed most of their theorizing between the first and second probes, generating roughly equal shares of accurate and erroneous theorizing. Between the next few theory probes, participants made far fewer modifications to their theories, although they shed a modicum of their erroneous theorizing while adding a measure of accurate thinking. In the last transition between theory probes, participants again started developing both accurate and inaccurate notions about how to approach their diagnoses.

In similar analyses, splitting theory development into its accurate and erroneous components showed that each followed a different temporal pattern of development (see Figure 5). Accurate theory development revealed significant linear, quadratic, and cubic components, using the same contrast weights as above, F = 21.77, 19.75, and 17.96, η_p^2 = .32, .30, .28, ps < .02, respectively. Although, for this analysis, it was the linear trend that proved the largest. For erroneous theory development, only the cubic trend was significant, $F(1, 44.4) = 17.40$, $p < .001$, $\eta_p^2 = .28$, explaining perhaps why overall theory development unfolded in a more cubic than linear fashion over experience.

Mediation

Our last set of analyses explored whether this time course of theory development explained the time courses of confidence and accuracy we saw. Our theory development measures were designed to give us "snapshots" of participant theories at six different points of the zombie task. We decided to take similar snapshots at those exact points in time for confidence and accuracy. Thus, for each point at which we collected theory measures, we also took confidence and accuracy data within three trials of that point. For example, for the first theory probe, we took data from the first three cases that participants encountered, for the next three theory probes, we took data from the three cases that preceded the probe and the three that followed it. For the last theory probe, we took data from cases 58 through 60. We adjusted individual confidence and accuracy data for any subject and patient profile effects, and then aggregated scores associated with each theory probe.[4] Thus, for each theory probe, we had confidence and accuracy data that represented participant's contemporaneous performance and

[4] One confidence judgment was omitted from subsequent analysis, lying 3.8 SDs away from its group mean, 1.9 SDs away from its nearest neighbor.

perception thereof. These data preserved the effects of experience we had seen previously. For accuracy, the linear component of improvement was preserved in an multi-level model (random-intercept, random-slope) weighted -5, -3, -1, 1, 3, and 5 for each time period, with participants as a random variable, $F(1, 187.5) = 11.50$, $p < .001$ $\eta_p^2 = .06$. For confidence, the cubic trend (weighted -5, 7, 4, -4, -7, 5) was similarly preserved, $F(1, 46.2) = 15.41$, $p < .001$, $\eta_p^2 = .25$.[5]

We then, first, looked to see if the cubic trend seen for confidence was explained by the fact that overall theory development followed the same cubic trend. This question reduces to a mediation analysis, looking to see if the cubic trend for confidence diminishes after theory development was controlled for. Above, we have demonstrated the cubic trend for both theory development and confidence. The only step remaining to establish mediation was to examine whether theory development remained correlated with confidence after the cubic trend is controlled for, whereas the cubic effect on confidence was reduced. Thus, we repeated the multi-level analysis on confidence with overall theory development as a covariate. Theory development continued to be significantly related to confidence, $b = .87$, $se_b = .12$, $p < .001$, $\eta_p^2 = .63$, with the cubic trend on confidence evaporating to nonsignificance, $F(1, 52.3) = .08$, *ns*, $\eta_p^2 = .01$, all consistent with mediation, Sobel $z = 4.29$, $p < .001$ (see Figure 7).

In a similar vein, mediational analyses showed that the linear trend in accurate theory development explained the linear trend in diagnostic accuracy in the zombie task. Above, we have documented the linear trend in both accurate theoretical development and diagnosis accuracy. To demonstrate mediation, we conducted a multi-level analysis on diagnostic accuracy including the linear trend and accurate theory development as predictors. Accurate theory

[5] We adopted this "snapshot" approach because our final objective was to see whether trends in confidence and accuracy were explained by the exact same trends in theory development, as embodied in our linear contrasts. That meant creating summary measures of confidence and accuracy that reasonably reflected participants' responses around the specific occasions we asked them to describe their theories and imposed our linear contrasts.

development still predicted diagnostic accuracy, $b = .33$, $se_b = .08$, $p < .001$, $\eta_p^2 = .06$. Some portion of the linear trend remained, $F(1, 276.4) = 4.23$, $p = .041$, $\eta_p^2 = .01$, but was significantly reduced, Sobel $z = 3.82$, $p = .002$.

Summary. In sum, Study 4 largely replicated the pattern we found in the first three studies of participant reactions as they gained experience with the zombie task. In addition, the study tied these patterns of confidence and accuracy to the underlying theorizing participants engaged in as they gained experience in the task. Participants displayed a burst of early theorizing that inflated confidence and created the beginner's bubble seen in the first three studies. After that bubble, participants settled into a pattern of theory incremental revision that increased accuracy but did not inflate confidence again until the very end of the task.

One aspect that the study did not replicate was an overall effect of overconfidence. We can speculate, however, that the methods we used in Study 4 dampened people's usual level of confidence. In stopping the medical diagnosis task and asking people to state their theories, we asked people not only to articulate what they "knew" about the task but also potentially confronted them with detailed knowledge they did not know or had doubts about. Recent work suggests that confronting people in such a way tends to lower their confidence (Hadar, Sood, & Fox, 2013; Walters, Fernbach, Fox, & Sloman, in press).

Study 5a and 5b: Financial Literacy

The studies so far have been laboratory-based. In Studies 5a and 5b we asked whether our results would generalize to a crucial skill in the general population, managing one's finances, focusing on data from the 2012 and 2015 panels of the Financial Industry Regulatory Authority (FINRA) survey on financial capability, conducted in partnership with the United States Department of the Treasury (Lin et al., 2016; Lusardi et al., 2013). Each panel queried a nationally representative sample of roughly 25,000 U. S. respondents on their financial history, habits, and opinions. Of key interest, each survey asked respondents to rate their "financial knowledge," and then presented them with a 5- (2012) or 6-item (2015) financial literacy test, querying their understanding of basic financial concepts such as inflation, compound interest, the relation between bond rates and prices, investment diversification, and risk.

Although it is a step away from the probabilistic learning tasks used in the lab studies presented herein, financial literacy is a multi-faceted task that serves as a particularly fitting domain to explore the development of perceived versus actual self-knowledge. Most under the age of 18 have little knowledge of personal finance (Avard, Manton, English, & Walker, 2005). Typically, until this age parents assume the responsibility of engaging in financial transactions for minors (Kramer, 1994, Cunningham, 2006, Schwartz, 2011). Teenagers cannot typically acquire credit cards and personal loans, purchase homes, or engage in many, if not most, financial transactions without adult supervision. Minors have limited financial abilities. Further, most primary and secondary educational systems do not teach financial literacy (Mandell, & Klein, 2009). It is therefore not unreasonable to assume that young adults are the least knowledgeable about finances and likely have very little knowledge on this topic.

We wished to see what happens to objective financial knowledge and subjective impressions of self-knowledge among young adults as they are thrust into the world and targeted by banks, credit cards, and the demands of independent life, typically with very little preparation to become consumers of financial products. As they grow older, they engage in more complicated financial transactions that should increase their knowledge. At times, these financial transactions provide rewards and other times financial mistakes are made. In addition, people receive informal advice from family members, friends, and the media about how to handle their money. That is, much like a probabilistic learning task, personal finance is a complex task that people learn via trial and error in a complicated, somewhat haphazard, information environment.

Thus, using data from the FINRA surveys, we made three predictions. First, financial literacy would incrementally increase with age. However, self-ratings of financial literacy across the lifespan will follow the same non-linear pattern observed in the lab: Confidence would surge as people began their adult years, then flatten out and potentially decrease across the middle years, only to rise once again as people approached their older years.

Method

Participants. Data were obtained from the National Financial Capability Study in 2012 (Study 5a) and 2015 (Study 5b) (Lin et al., 2016; Lusardi et al., 2013). These data represent a nationally representative sample of American adults with at least 500 respondents from each U. S. state. The datasets, already stripped of participant identity, are publically available from the FINRA website (http://www.usfinancialcapability.org/downloads.php). The total sample size was 25,509 in 2012 and 27,564 in 2015.

Procedure. The survey comprises a comprehensive questionnaire on basic demographics, financial history, money habits, and financial opinions. As part of the survey, participants answered five multiple-choice questions in 2012 or six in 2015 to assess their financial literacy (e.g., "If interest rates rise, what will typically happen to bond prices?"). Participants were further asked to assess their self-perceived financial knowledge on a seven-point scale ranging from 1(very low) to 7 (very high) before they encountered the financial literacy quiz.

The survey aggregates respondents into 6 age groups (i.e., 18-24 years of age, 25-34, 35-44, 45-54, 55-64, 65 plus). It also records participant gender (which we coded female = 1, male =

2), education level across six categories (i.e., did not complete high school, high school diploma (regular or GED), some college, associate's degree, bachelor's degree, post-graduate degree), and yearly income (i.e., less than \$15,000, less than \$25,000, less than \$50,000, less than \$75,000, less than \$100,000, less than \$150,000, \$150,00 or more).

Results and Discussion

Only participants who reported their age, self-perceptions of their financial knowledge, and the financial literacy quiz were included in these analyses. The final sample consisted of 24,814 Americans in 2012 and 25,901 in 2015. In all analyses reported below, we weighted respondents' data according to weights provided in the FINRA datasets to achieve a representative portrait of the United States.

Actual Financial Literacy. We subjected scores of on the financial literacy test (depicted in Table 2) to two separate ANOVA analyses. In Model 1, we examined the relationship of age to literacy, examining across our six age groups the strength of the linear trend (weighting groups -5, -3, -1, 1, 3, 5, from 18-24 age group to the 65 plus age group, respectively), quadratic trend (weights were 5, -1, -4, -4, -1, 5), and cubic trend (weights were -5, 7, 4, -4, -7, 5). As such, we had tests of each trend that were independent of each other. This model showed that all three trends were significant in both 2012 and 2015 panels (see Table 3), except for the cubic trend in 2015, but that the linear trend was much stronger than the other two. In fact, the linear trend explained 93% and 98% of the between-group variance due to age in both panels.

Model 2, the second analysis, added covariates for education, income level, and gender. Education and income proved to have a positive relationship with literacy; in addition, men outscored women in both surveys (see Table 3). That said, the strong linear trend due to age emerged once again, explaining over 91% and nearly 94% of the between-group variance attributable to age in the 2012 and 2015 panels, respectively. The coefficients for quadratic and cubic trends flipped in sign or became nonsignificant in both panels, suggesting that these trends were not reliable.

Perceived Financial Literacy. We subjected self-ratings of financial knowledge to three different regression analyses (see Table 4). In the first, Model 1, we regressed self-perceptions of knowledge onto linear, quadratic, and cubic trends according to age, using the same group weights as above. As seen in Table 4, all three trends were significant. Of key interest, the cubic trend explained 17% and 24% of the between-group variation due to age in the 2012 and 2015 panels, respectively.

In our second analysis, Model 2, we again looked for linear, quadratic, and cubic trends, this time controlling for actual financial literacy. All three trends emerged, with the cubic trend explaining 32% and 22% of the between-group variance attributable to age in the 2012 and 2015 panels, respectively. Self-perceived literacy correlated with actual literacy at only a modest level, $r(24{,}812) = .25$ and $r(26{,}899) = .21$, for 2012 and 2015 panels, respectively, ps < .001. Figure 8 depicts the self-rating given as a function of age, for both raw analysis (Model 1, see left panel) and one controlling for actual knowledge (Model 2, see right panel). Self-ratings surged between the youngest age group and the one aged 18-24 years. They then flattened or declined up to the group aged 45-54 years old, after which self-ratings of financial knowledge rise again. The pattern was more pronounced after controlling for actual financial knowledge.

Finally, in Model 3, we added education, income, and gender to the regression analysis. Education and income were both associated in either survey with enhanced self-ratings of skill. Men also rated themselves as more skilled than women. Beyond this, all three age trends continued to be significant predictors of self-rated knowledge, with the cubic trend still explaining 16% and 21%, for 2012 and 2015 panels, respectively, of the between-group variance attributable to age.

Summary. In sum, in Study 5a and 5b we found over the life-course a picture of confidence and objective skill that resembled what we found over the short-term in the lab. Confidence and skill do not rise in tandem. Confidence appears to outstrip learning in the early stages of adulthood, only for learning to catch up, if it does at all, slowly over more experience.

General Discussion

Søren Kierkegaard once famously observed that although life must be lived forwards, it could be understood only backwards. Thus, beginners, those with the most life left to live, are often the ones least prepared to make decisions with proper certainty about how to live it, in that those with little understanding tend to be the most overconfident in what they decide (Dunning, 2011; Dunning, et al., 2003; Kruger & Dunning, 1999).

As such, we explored overconfidence among beginners to trace how it may develop. As we expected, we found that people do not begin harboring overconfidence, but it takes only a little experience to prompt them toward that overinflated confidence. Beginners quickly develop a bubble of overconfidence that begins to flatten or deflate only after a while. It does take a little learning for this overconfidence to develop.

We documented this beginner's bubble in our first three studies, in which we confronted participants with multi-cue probabilistic learning tasks. As our participants gained experience and feedback with the tasks, their accuracy rose in an incremental and linear fashion. The confidence they expressed, however, was anything but linear. Across the studies, participants showed no overconfidence as they began, but after 9 to 14 learning trials their confidence rose well beyond where their accuracy lay. However, that confidence soon leveled off as accuracy continued its steady rise. Accuracy never matched confidence, however. In an unexpected finding, we discovered that confidence began another increasing trend after a pause that ultimately kept people roughly at a constant level of overconfidence as they ended the task.[6]

In Study 4, we assessed a psychological mechanism, exuberant theorizing, that we asserted was potentially responsible for this beginner's bubble. We predicted that people would rapidly form theories about how to approach the tasks we confronted them with, but that their theorizing would far outstrip the validity of the small amount of data they based it on. This is exactly what we found. Within 12 trials of experience in diagnosing zombie illnesses, participants held confident theories about which symptoms predicted zombie illness.[7] Although roughly 63% of the notions in their theories were wrong, these theories produced confidence—and overconfidence—in diagnosis. Fortunately, with further experience, participants revised those theories in an accurate direction, ultimately achieving roughly 46% accuracy in their theory, and so continued to achieve incrementally better performance without an overall appreciable rise in confidence, until the very end of the experimental session. As such, mediational analyses on theory development successfully accounted for the initial beginner's bubble we observed in confidence.

On Expertise

[6] Discerning readers may worry that although we observed a significant cubic trend in the aggregate, it is possible that no participant displayed it at the individual level. It is only in averaging trends across participants that the cubic trend arises. We discuss this issue in supplemental materials. According to our coding scheme, when we classified individual participants according to the specific trend in confidence each displayed, we find that 49% displayed a positive cubic trend, with an additional 23% displaying primarily a negative quadratic trend. Both trends, which characterize 72% of participants across studies, are consistent with a "beginner's bubble" pattern of rapidly inflating confidence.

[7] We also note that participants were so exuberant in their theorizing that 40 out of 47 in Study 4 expressed some partial theory of zombie illness even before they saw their first patient, clearly applying inferences from their world knowledge (see also Fischhoff & Slovic, 1980, for similar behavior).

Participant theorizing also appeared to explain the unexpected "bonus" finding revealed in each of our studies, an uptick in confidence that emerged toward the end of our experimental sessions. More specifically, analyses showed that participants became more confident in their theories, both accurate and inaccurate elements, as they neared the end of the experimental session. This increase in theoretical development accounted for the unexpected tail-end increase in confidence.[8] This pattern meant that although accuracy continued to rise in a linear fashion as participants gained experience, confidence resumed its rise in such a way to insure that participants would always retain a significant level of overconfidence.

We believe, however, that this tail-end rise in confidence is worthy of further study. It suggests, as has been frequently been found in the literature, that experts can be just as prone to overconfidence as nonexperts are, despite greater accuracy (Hazard & Peterson, 1973; Cambridge & Shreckengost, 1978; McKenzie, Liersch, & Yaniv, 2008; Oskamp, 1962; Von Holstein, 1972; Moore, 1977; Neale & Bazerman, 1990; Wagenaar & Keren, 1986, Hynes & Vanmarcke, 1976, although see contrary evidence in Ericcson & Smith, 1991; Wallsten and Budescu, 1983). Our data provide a speculative explanation for overconfidence among experts, one that should be tested more formally in future research. As participants become more experienced, they develop more accurate components in the theories they use for judgment, but they also continue to possess hefty components of error in their theory as well. Indeed, in Study 4, on average 54% of participants' theoretical development score represented erroneous notions rather than accurate ones. Possessing substantial "knowledge" that is false may be enough to

[8] A mediational analysis supports this. If we look at the last three theory probes, representing the last-minute rise with weights (-1, -1, +2), we see a significant trend in both theory development, $F(1, 45.2) = 12.44$, $p = .001$, $\eta_p^2 = .22$, and confidence, $F(1, 45.3) = 21.57$, $p < .001$, $\eta_p^2 = .33$. If we control for confidence, total theory development continues to predict confidence, $b = .31$, $p = .044$, $\eta_p^2 = .31$, with the time trend in confidence is reduced to a marginally significant degree, Sobel $z = 1.91$, $p = .056$, two-tailed.

produce inflated confidence even among experts.

Real World Echoes

Finally, in Study 5a and 5b, we explored the generality of the beginner's bubble by moving away from probabilistic learning to the no less complex task of managing one's personal finances. Using data from two different panels of the National Financial Capability Survey, each involving more than 25,000 respondents, we showed that self-perceptions of financial expertise followed a beginner's bubble pattern across the life course, with young adults (aged 25-34 years old) more confident than those younger (18-24), but no less confident than those older than them (i.e., 35-44, 45-54). This youthful bubble of confident self-perception arose even though, again, actual financial literacy rose only incrementally and slowly throughout the lifespan. And, again mimicking our lab data, older adults again started to grow more confident in their financial knowledge after a pause in middle age.

Of course, we must place a caveat on our interpretation of the findings. We attribute differences across the groups in these studies to participant age, but the data are cross-sectional. The best way to implicate age and experience in any dissociation between confidence and accuracy would be a longitudinal analysis tracking the same respondents through time. We hope future research will be able to fulfill this goal.

Relation to the Dunning-Kruger Effect

The results presented here also have implications for the Dunning-Kruger effect, the fact that poor performers tend not to know how poorly they perform, thus exhibiting marked degrees of overconfidence (Dunning, 2011; Dunning et al., 2003; Ehrlinger et al., 2008; Kruger & Dunning, 1999). Here, we show a circumstance in which people are clearly aware that they are poor performers—namely, rank beginners. Our participants just starting the task were well-calibrated about how meager their accuracy would be, although they shed that calibration in short due course.

These data, however, do suggest a boundary condition for the Dunning-Kruger effect. The individual has to pass some minimal threshold of learning or experience before they begin to show the outsized confidence associated with poor performers. To be sure, Kruger and Dunning (1999) in their initial discussion of the effect noted that such boundary conditions might exist. The present data help to specify one important circumstance that can serve as a boundary, namely, whether a person is an absolute beginner at a task or skill. For these individuals, the Dunning-Kruger effect may not apply. However, a little experience might pass them into a circumstance in which they become some of the most vulnerable individuals to the Dunning-Kruger effect.

Questions for Future Research

Taken as a whole, our results present a programmatic and replicable pattern of overconfidence among beginners. That said, we hasten to add that this work must stand as only a first comment on the issue. There are many aspects of learning that may change or augment the conclusions we reach here—and these aspects are worthwhile candidates for further research.

Probabilistic Learning vs. Memory. Indeed, our work already presents an apparent contradiction to a well-studied phenomenon already in the literature, the underconfidence with practice (UWP) effect, which occurs in studies of memory (Koriat, 1993, 1995, 1997). In these studies, participants memorize lists of word pairs and then complete a cued recall task in which they are presented with a word and are asked to recall the other word paired with it. In a first round of this task, reminiscent of the initial patterns of confidence reported here, participants tend to be well-calibrated in their confidence about their memory performance (Koriat, 1993).

They then restudy the list and complete a round of the second recall task. Their recall performance rises but, in apparent contradiction to our results, their confidence fails to rise, with participants thus displaying clear underconfidence.

How can we reconcile our pattern of overconfidence with the typical pattern of underconfidence seen in the UWP effect? We think there are two possible reconciliations. The first is to observe that we presented participants with a rather novel task, either diagnosing illness or determining which lie detector works best, but that research on the UWP effect focuses on a task, memory, that participants are already familiar with. As students, they are well-acquainted with memorizing material that they will later report depending on what cues are present in a question. As such, the UWP does not contain the crucial circumstance, confronting a completely novel task, that we studied here.

However, if one stipulates that our tasks and the recall tasks used in UWP studies are both novel tasks, the other route to reconcile their divergent results is to note each contains different tasks involving different cognitive demands, processes, and influences (Hoffman, von Helversen, & Rieskamp, 2014). In our probabilistic learning tasks, participants could use feedback to construct explicit strategies about how to succeed in making their predictions. As such, feedback on one trial could lead to revised or consolidated strategies to apply to the next trial, leading to a rise in confidence as participants honed their explicit theories of how to approach the task. It appears that in UWP studies, participants also hone theories of what they can remember, in that they become more accurate distinguishing words they will remember from those they will forget (Koriat, 1997), but they miss the overall impact of one singular influence, one operating under the radar of conscious awareness, that improves their overall memory performance across learning sessions. This influence is the implicit, nonconscious and beneficial effect that follows from repetition in the study of the same words. However, as not part of their conscious strategy, the impact of repetition lies outside of what they may think about when forming confidence estimates (Koriat, 1993, 1997).

As such, future work on beginner's confidence will have to explore different tasks, subject to different demands and influences on both confidence and accuracy, to determine the generality of the beginner's bubble finding we observed here. That, however, is not the only open question for future work to consider.

Unknown Task Variations. Further, exuberant theorizing may not be the only mechanism that produces a beginner's bubble of overconfidence. In our studies, we presented participants with a largely constrained and repeatable task. Other tasks might produce initial overconfidence because they are less well-defined and present a wider range of circumstances and demands. Participants may not have the chance to encounter rarer variations or more challenging complications of a task until they are far into their experience, with these new experiences catching people as overconfident.

For example, people learning to fly airplanes may not encounter all the conditions that can make flying risky (e.g., adverse weather conditions, equipment failure, navigational issues) until they are well into the "killing zone" that aviators worry about. Surgeons, too, may have the chance to encounter rare but problematic situations that complicate surgery only after accumulating a good deal of experience (e.g., a rare parasite, oddly placed blood vessels, missing or misshapen organs). If they were unlucky enough to encounter these tricky situations as relative beginners, they may approach them with ample caution and underconfidence. However, if they encounter them only after gaining a good deal of experience, that experience may lead them to believe they can deal with these challenging situations. For example, a more experienced pilot may dismiss the chance of inclement weather complicating his or her flight plan whereas a beginner will immediately seek to land the airplane. In short, the task may contain a number of "unknown unknown" challenges, unexpected and relatively rare task complications, that take a while to reveal themselves. As beginners, they would react to these challenges with humility. As more experienced experts, they may find themselves with the need to reacquaint themselves with that humility.

Censored or Contaminated Feedback. Further complicating the picture is the type of feedback people may acquire as they learn. Herein, we tested participants in an ideal situation: They received feedback after each and every trial and could take all the time they wanted to view and mull over that feedback before turning to the next decision. In life, feedback is often constrained, censored, or unavailable (Denrell, 2005; Einhorn & Hogarth, 1978; Fetchenhauer & Dunning, 2010; Fischer & Budescu, 2005). Company heads receive feedback only about those individuals they hire, not the ones they turn away. Doctors may find out the fate only of those patients they admit for illness, but not those they send away. People gain social feedback about the people they trust, but not those they distrust. Past work suggests that such asymmetric feedback prompts people toward overconfidence (Fischer & Budescu, 2005; Smillie, Quek, & Dalgleish, 2014) and fails to correct people's preconceived ideas that happen to be mistaken (Elwin, 2013; Fetchenhauer & Dunning, 2010). As such, situations of selective or incomplete feedback may lengthen the beginner's bubble of overconfidence we saw here.

Moreover, people may act upon their judgments in ways that biases the reactions of other people, leading to a contamination of feedback received (Einhorn & Hogarth, 1978). For example, deciding that another person will be aggressive might cause the social perceiver to make pre-emptive aggressive moves. Once made, these moves impel the other person to act in kind, even though he or she may have had an initial preference to act in a more prosocial way and would have done so without the aggressive provocation (Kelley & Stahelski, 1970a, 1970b). Because they contaminate the feedback they receive, people may fail to correct the overconfidence they acquire while beginners.

Self-Selection. Another dimension that might influence the relation of confidence and accuracy would be the by which why people begin the task in the first place. Herein, we assigned participants, for example, to the zombie task. What would happen if participants instead had a chance themselves to choose their tasks, such as people choose careers and hobbies in the real world. We can presume that people, if given the freedom, would tend to choose tasks that they consider themselves talented in while avoiding those at which they think they are lackluster. Past research bears this intuition up. People tend to volunteer for tasks in which they already are confident they will do well (Camerer & Lovallo, 1999; Koellinger, Minniti, & Schade, 2007). As such, people who volunteer for tasks, relative to those who are assigned to the task, may show more initial overconfidence, or grow into their overconfidence much more quickly.

The Shape of Learning. Finally, the shape of learning may influence whether a beginner's bubble develops. Not all probabilistic tasks necessarily follow a linear function of learning (Newell & Rosenbloom, 1981; Ritter & Schooler, 2002). What if the task is easy, or a matter of flash of insight? What if people fail to learn anything? Each variation in learning may result in different patterns in confidence or overconfidence (Gottlieb et al, 2013).

Further, we should note that further work should determine whether our results are the effects of a "little learning" or just raw experience that requires no learning. Others have shown that mere access to information increases feelings of knowing. Merely being exposed to information increases confidence even when that information even is false or irrelevant (Koriat, 1993,1995; Gill, Swan, & Silvera, 1998; Fiedler et al., 1996; Oskamp, 1965; Tsai, Klayman, & Hastie, 2008). As such, the curvilinear pattern in confidence we observed may occur in some probabilistic tasks even in the absence of learning.

Concluding Remarks

In sum, this research suggests that learning leads to the development of overconfidence, but that we as a research community should not be too confident that we know all the nuances of that relationship. Herein, we have begun a conversation by proposing that beginners never having performed a task (i.e., the truly incompetent) are often quite aware of their inability. With a little learning, however, beginners quickly come to believe they know much if not all there is to know. They formulate faulty but forcefully-held theories about how to approach tasks based on the shards of experience they have gained.

As such, it takes just a little learning for people to overrate their abilities. This leaves learners with a dilemma. On the one hand, learning is necessary to acquire abilities. On the other hand, this same learning, at least for a time, leads people to overestimate those abilities inappropriately. A potential resolution to this dilemma might require being mindful of English philosopher R. G. Collingwood when he observed that people cease to be beginners in any craft or science, and become instead masters, at the moment they realize they are going to be beginners for the rest of their lives.

References

Adams, L. (2017). Learning a skill is easier said than done. Retrieved from: http://www.gordontraining.com/free-workplace-articles/learning-a-new-skill-is-easier-said-than-done/#

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, *103*, 718-735.

Armor, D. A., & Taylor, S. E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology*, *30*, 309-379.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*, 258-290.

Avard, S., Manton, E., English, D., & Walker, J. (2005). The financial knowledge of college freshmen. *College Student Journal*, *39*, 321-339.

Avon, S. L., Victor, C., Mayhall, J. T., & Wood, R. E. (2010). Error rates in bite mark analysis in an in vivo animal model. *Forensic Science International*, *201*, 45-55.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*, 773-806.

Becker, G.M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research*, *9*, 226-232.

Benjamin, D. J., Rabin, M., & Raymond, C. (2015). A model of nonbelief in the law of large numbers. *Journal of the European Economic Association*, *14*, 515-544.

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*, S2-S23.

Billeter, D., Kalra, A., & Loewenstein, G. (2010). Underpredicting learning after initial experience with a product. *Journal of Consumer Research*, *37*(5), 723-736.

Blavatskyy, P. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty*, *38,* 39–49.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological review*, *50*(3), 255-272.

Cambridge, R. M., & Shreckengost, R. C. (1978). *Are you sure?: The subjective probability assessment test*. Information Science Center, Office of Training, Central Intelligence Agency.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, *89,* 306-318.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*, 271-280.

Craig, P. (2013). *The killing zone: How & why pilots die*. New York: McGraw Hill Professional.

Cunningham, L. (2006). Question of capacity: Towards comprehensive and consistent vision of children and their status under law. *UC Davis Journal of Juvenile Law Policy, 10,* 275-378.

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112,* 951-978.

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology*, (pp. 247-296). New York: Elsevier.

Dunning, D. (2012). *Self-insight: Roadblocks and detours on the path to knowing thyself*. Psychology Press.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58*, 568-581.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69-106.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83-87.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware? Further explorations of (lack of) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105,* 98-121.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological review*, *85*, 395-416.

Elwin, E. (2013). Living and learning: Reproducing beliefs in selective experience. *Journal of Behavioral Decision Making*, *26*, 327-337.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits.* Cambridge University Press.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*, 37–64.

Fetchenhauer, D. & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism in the trustworthiness of others. *Psychological Science, 21*, 189-193.

Fiedler, K., Walther, E., Armbruster, T., Fay, D., & Naumann, U. (1996). Do YouReallyKnow what you have seen? intrusion errors and presuppositions effects on constructive memory. *Journal of Experimental Social Psychology*, *32*(5), 484-511.

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes*, *98*, 39-53.

Fischoff, B., & Slovic, P. (1980). A little learning ...: Confidence in multicue judgment tasks. In R. Nickerson (Ed.), *Attention and performance, VIII* (pp. 779-800). Hillsdale, NJ: Erlbaum.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, *3*, 552-564.

Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G. C., Wolf, F. M., Heckerling, P. S., Fine, P.L., Miller, T.M, & Elstein, A. S. (2005). Do physicians know when their diagnoses are correct?. *Journal of General Internal Medicine*, *20*, 334-339.

Gill, M. J., Swann Jr, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, *75*, 1101-1114.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585-593.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411-435.

Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology*, *18*, 186-221.

Hadar, L., Sood, S., & Fox, C. R. (2013). Subjective knowledge in consumer financial decisions.

Journal of Marketing Research, *50*, 303-316.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*, 385-410.

Hayward, M. L., & Hambrick, D. C. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly*, *42*, 103-127.

Hayward, M. L., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science*, *52*, 160-172.

Hazard, T. H., & Peterson, C. R. (1973). *Odds versus probabilities for categorical events* (No. TR-73-2). McLean VA.: Decisions and Designs Inc.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243-259.

Hoffman, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*, 2242-2261.

Hynes, M. E., & Vanmarcke, E. H. (1976). *Reliability of embankment performance predictions*. In: Proceedings of the ASCE Engineering Mechanics Division Specialty Conference, Waterloo (pp. 367-383). Waterloo, Ontario: University of Waterloo Press.

Johnson, D. D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions.* Cambridge, MA: Harvard University Press.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. (1968). Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology*, *10*, 317.

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

Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior, 1967: Aversive stimulation* (pp. 9-31). Coral Gables, FL: University of Miami Press.

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality, 18*, 431-439.

Kelley, H. H., & Stahelski, A. J. (1970a). The inference of intentions from moves in the Prisoner's Dilemma game. *Journal of Experimental Social Psychology, 6*(4), 401-419.

Kelley, H. H., & Stahelski, A. J. (1970b). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of personality and social psychology, 16*(1), 66.

Knecht, W. R. (2013). The “killing zone” revisited: Serial nonlinearities predict general aviation accident rates from pilot total flight hours. *Accident Analysis & Prevention, 60*, 50-56.

Koellinger, P., Minniti, M., & Schade, C. (2007). “I think I can, I think I can”: Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology, 28*, 502-527.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological review, 100*(4), 609-639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*(3), 311-333.

Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Kramer, D. T. (1994). *Legal rights of children* (2nd ed.). Minneapolis, MN: Thomson/West.

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*, 1121-1134.

Lin, J. T., Bumcrot, C., Ulicny, T., Lusardi, A., Mottola, G., Kieffer, C., & Walsh, G. (2016). *Financial capability in the United States 2016*. Washington, DC: FINRA Investor Education Foundation.

Little, D. R., & Lewandowsky, S. (2012). Multiple-cue probability learning. In *Encyclopedia of the Sciences of Learning* (pp. 2386-2388). Springer US.

Lusardi, A., Bumcrot, C., Lin, J., & Ulicny, T. (2013). *Financial capability in the United States: Report on findings from the 2012 National Financial Capability Study.* Washington, DC: FINRA Investor Education Foundation.

Mandell, L., & Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning, 20,* 15-24.

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making*, *17*, 281-295.

McKenzie, C. R. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 771-792.

McKenzie, C. R., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you?. *Organizational Behavior and Human Decision Processes*, *107*(2), 179-191.

Moore, P. G. (1977). The manager's struggle with uncertainty. *Journal of the Royal Statistical Society*, *140*, 129-165.

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

Neale, M. A., & Bazerman, M. H. (1990). *Cognition and rationality in negotiation*. New York: Free Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, *1*, 1-55.

O'Callaghan, E., Miyamoto, A., Takahashi, H., & Fujita, R. (1990). Do beginner endoscopists make erroneous diagnosis? *The Showa University Journal of Medical Sciences*, *2*, 121-125.

Ono, K. (1987). Superstitious behavior in humans. *Journal of the Experimental Analysis of Behavior*, *47*, 261-271.

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs: General and Applied*, *76*, 1-27.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*, 261-265.

Pateros, J. (2017). The four stages of learning. Retrieved from: https://processcoaching.com/fourstages.html.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, *46*(4), 657-685.

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin*, *33*, 1492-1502.

Ritter, F. E., & Schooler, L. J. (2002). The learning curve in International encyclopaedia of the social and behavioural sciences. 8602-8605.

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review*, *33*, 7-17.

Schatlo, B., Martinez, R., Alaid, A., von Eckardstein, K., Akhavan-Sigari, R., Hahn, A., Stockhammer, F., & Rohde, V. (2015). Unskilled unawareness and the learning curve in robotic spine surgery. *Acta Neurochirurgica*, *157*, 1819-1823.

Schotter, A., & Trevino, I. (2014). Belief elicitation in the laboratory. *Annual Review of Economics, 6*, 103-128.

Schwartz, A. A. (2011). Old Enough to Fight, Old Enough to Swipe: A Critique of the Infancy Rule in the Federal Credit CARD Act. *Utah Law Review*, 407-432.

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1003-1021.

Smillie, L. D., Quek, B. K., & Dalgleish, L. I. (2014). The impact of asymmetric partial feedback on response bias. *Journal of Behavioral Decision Making*, *27*, 157-169.

Swann, W. B., Jr. & Gill, M.J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology, 73*, 747-757.

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, *51*, 176-197.

Trautmann, S. T., & van de Kuilen, G. (2015). Belief elicitation: A horse race among truth serums. *The Economic Journal*, *125*, 2116-2135.

Tsai, C. I., Klayman, J., & Hastie, R. (2008). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes*, *107*, 97-105.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. .*Psychological Bulletin*, *76*, 105-110.

Vallone, R. P., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58*, 582-592.

Von Holstein, C. A. S. S. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance*, *8*, 139-158.

Wagenaar, W. A., & Keren, G. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Maneini, and D. D. Woods (Eds.), *Intelligent decision support in process environments* (pp. 87-107). Berlin: Springer.

Wallsten, T. S., & Budescu, D. V. (1983). State of the art—encoding subjective probabilities: A psychological and psychometric review. *Management Science*, *29*, 151-173.

Walters, D. J., Fernbach, P. M., Fox, C. R., & Sloman, S. A. (in press). Known unknowns: A critical determinant of confidence and calibration. *Management Science.*

Wikipedia. (2017). The four stages of competence. https://en.wikipedia.org/wiki/Four_stages_of_competence. Accessed February 27, 2017.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*, 1006-1014.

Yarritu, I., Matute, H., & Luque, D. (2015). The dark side of cognitive illusions: When an illusory belief interferes with the acquisition of evidence-based knowledge. *British Journal of Psychology*, *106*, 597-608.

Table 1. Linear, Quadratic, and Cubic Models Predicting Confidence from Experience in Studies 1 – 4.

	Measure				
Model	*b*	se_b	*p*	η_p^2	*BIC*
		Study 1			
Linear	.29	.045	< .001	.52	19011
Quadratic					18907
Linear	.29	.045	< .001	.52	
Quadratic	-.01	.002	< .001	.49	
Cubic					18889
Linear	.06	.057	*ns*	.01	
Quadratic	-.01	.002	< .001	.49	
Cubic	.0004	.00008	< .001	.22	
		Study 2			
Linear	.18	.034	< .001	.36	23929
Quadratic					23838
Linear	.18	.034	< .001	.36	
Quadratic	-.007	.0017	< .001	.28	
Cubic					23810
Linear	-.05	.06	*ns*	.01	
Quadratic	-.007	.0017	< .001	.28	
Cubic	.0004	.00008	< .001	.44	
		Study 3			
Linear	.30	.04	< .001	.54	24506
Quadratic					24355
Linear	.30	.04	< .001	.54	
Quadratic	-.01	.002	< .001	.41	
Cubic					24289
Linear	.13	.08	*ns*	.05	
Quadratic	-.01	.002	< .001	.41	
Cubic	.0003	.0001	.004	.15	
		Study 4			
Linear	.27	.046	< .001	.43	23455
Quadratic					23354
Linear	.27	.046	< .001	.43	
Quadratic	.006	.002	.012	.13	
Cubic					23348
Linear	.01	.07	*ns*	.04	
Quadratic	-.006	.002	.012	.13	
Cubic	.0003	.00009	.001	.20	

Table 2. Actual Performance on Financial Literacy Test as a Function of Age.

	Age Group					
Panel Year	18-24	25-34	35-44	45-54	55-64	65 plus
2012						
Right	42.6 (.50)	51.7 (.41)	57.8 (.43)	62.0 (.39)	65.0 (.41)	69.8 (.44)
Wrong	22.4 (.35)	20.3 (.28)	16.7 (.30)	14.6 (.27)	13.9 (.28)	12.1 (.30)
IDK	32.4 (.49)	27.4 (.40)	25.0 (.40)	23.2 (.40)	20.6 (.24)	17.4 (.43)
n	2436	4125	4148	5073	4725	4194
2015						
Right	40.6 (.46)	46.2 (.38)	52.5 (.40)	55.8 (.40)	58.8 (.39)	62.4 (.37)
Wrong	25.6 (.33)	25.2 (.28)	21.8 (.29)	19.7 (.28)	18.4 (.28)	17.1 (.28)
IDK	33.3 (.47)	28.2 (.39)	25.3 (.41)	23.8 (.39)	22.0 (.39)	19.7 (.39)
n	2952	4887	4470	4902	4718	4992

Note: Scores on the test are expressed in terms of percents. Figures in parentheses are standard errors. Wrong = wrong answer chosen IDK = responded “I don’t know”.

*$p < .05$ **$p < .01$ ***$p < .001$

Table 3. Age Trends in Actual Financial Literacy (Studies 5a and 5b).

	Model 1				Model 2			
Measure	*b*	*F*	*p*	η^2	*b*	*F*	*p*	η^2
				Study 5a: 2012 Panel				
Age Trend								
Linear	.13	2232.27	<.001	.080	.100	1599.76	<.001	.061
Quadratic	-.02	68.07	<.001	.002	-.000	.02	*ns*	.000
Cubic	.007	20.73	<.001	.001	-.003	5.50	.019	.0002
Education					.295	1345.03	<.001	.052
Income					.143	1154.29	<.001	.045
Gender					.238	898.95	<.001	.035
				Study 5b: 2015 Panel				
Age Trend								
Linear	.117	1860.97	<.001	.065	.094	1393.28	<.001	.049
Quadratic	-.014	26.40	<.001	.001	.017	48.31	<.001	.002
Cubic	.012	42.73	<.001	.002	-.005	9.77	.002	.0003
Education					.258	1351.98	<.001	.048
Income					.165	1166.16	<.001	.042
Gender					.261	845.44	<.001	.030

Table 4. Age Trends in Perceived Financial Literacy (Studies 5a and 5b).

	Model 1				Model 2				Model 3			
Measure	*b*	*F*	*p*	η^2	*b*	*F*	*p*		*b*	*F*	*p*	η^2
					Study 5a: 2012 Panel							
Age Trend												
Linear	.056	486.94	<.001	.02	.031	143.56	<.001	.006	.026	102.47	<.001	.004
Quadratic	.009	15.77	<.001	.001	.013	33.58	<.001	.001	.021	88.45	<.001	.004
Cubic	.015	101.17	<.001	.003	.014	86.96	<.001	.004	.009	42.00	<.001	.002
Actual Literacy					.194	1106.9	<.001	.040	.114	325.22	<.001	.013
Education									.095	132.96	<.001	.005
Income									.092	455.93	<.001	.018
Gender									.079	96.12	<.001	.004
					Study 5b: 2015 Panel							
Age Trend												
Linear	.045	405.73	<.001	.015	.027	138.48	<.001	.005	.021	96.61	<.001	.004
Quadratic	.004	3.53	.060	.0001	.006	9.41	.002	.0003	.020	90.70	<.001	.003
Cubic	.015	125.85	<.001	.005	.015	123.74	<.001	.005	.011	51.04	<.001	.002
Actual Literacy					.143	979.50	<.001	.035	.082	284.26	<.001	.010
Education									.047	66.21	<.001	.003
Income									.102	661.35	<.001	.024
Gender									.081	123.34	<.001	.005

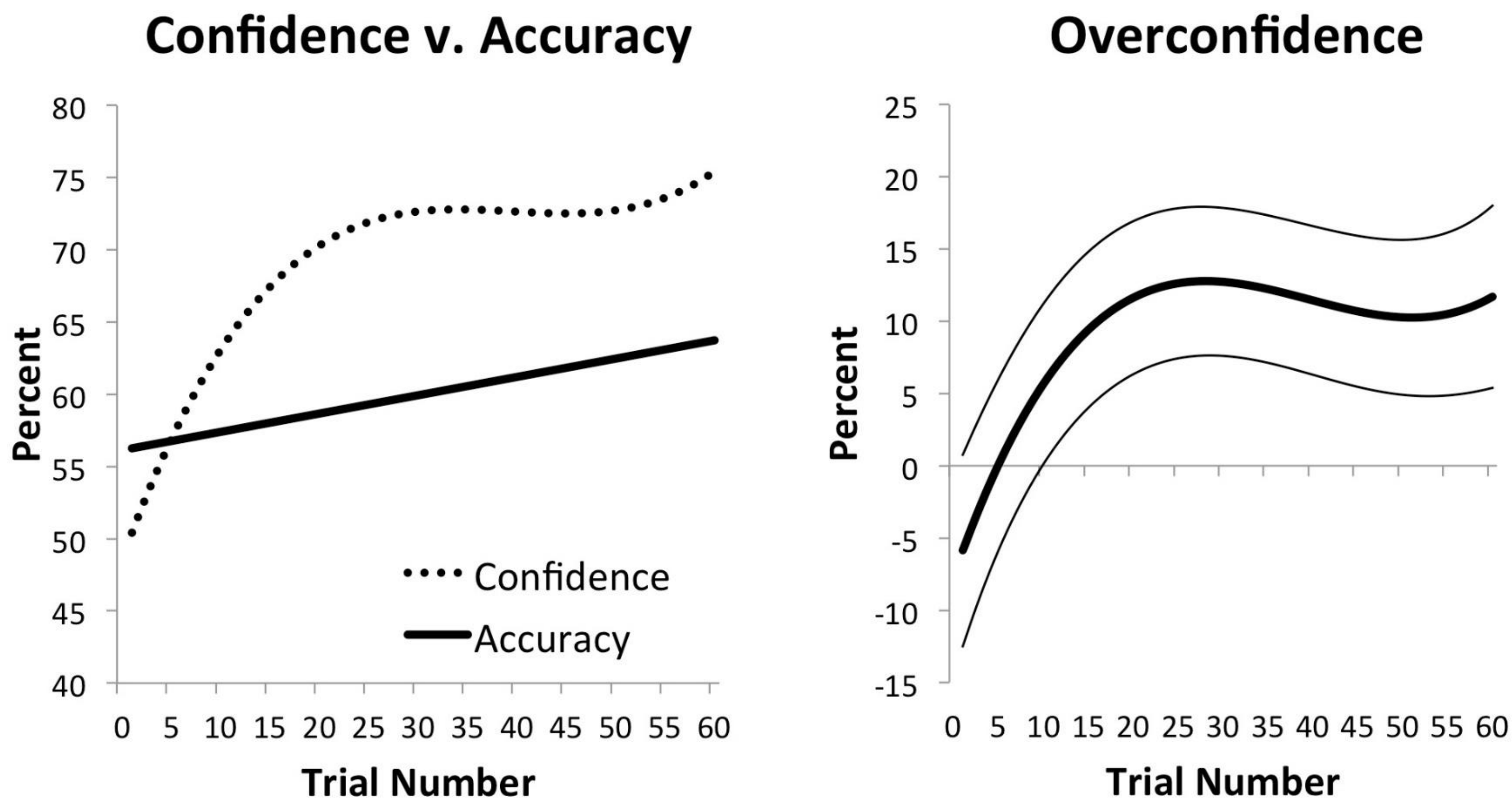


Figure 1. Confidence and accuracy trends over 60 diagnosis trials (Study 1). . Left Panel: Confidence and accuracy trends. . Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

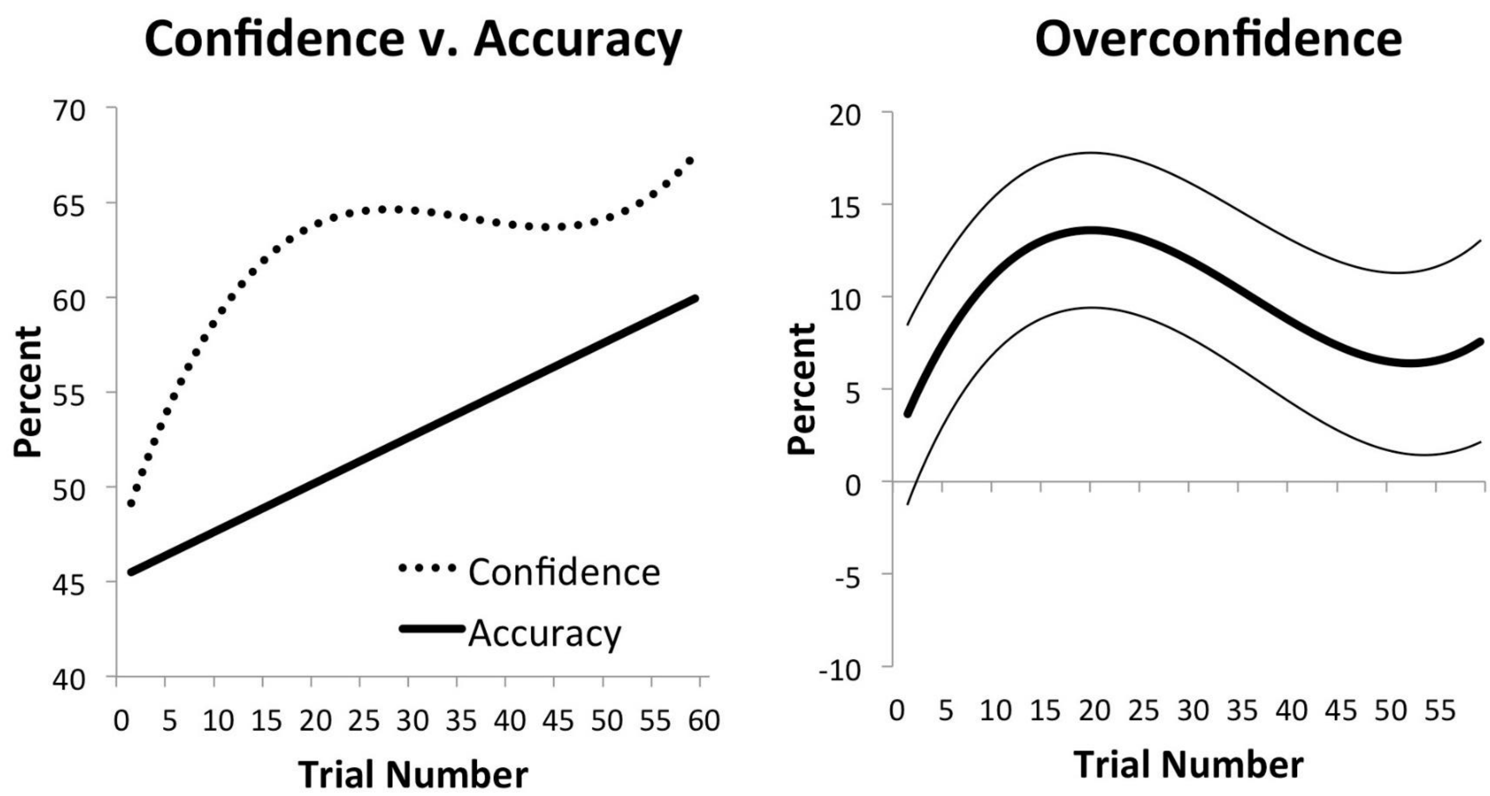


Figure 2. Confidence and accuracy trends over 60 lie detection trials (Study 2). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

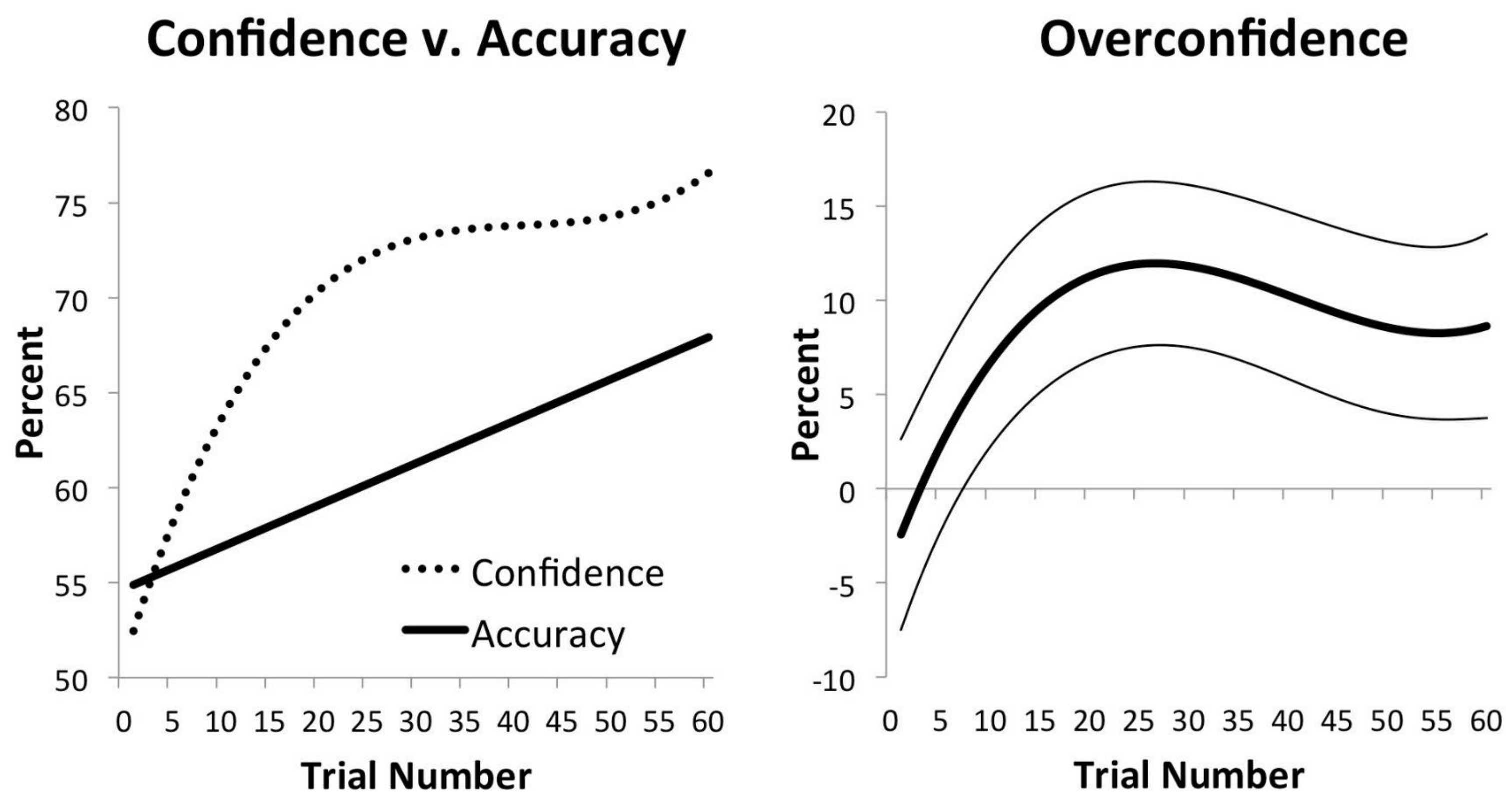


Figure 3. Confidence and accuracy trends over 60 diagnosis trials (Study 3). . Left Panel: Confidence and accuracy trends. . Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. . Upper and lower lines represent 95% confidence interval for the trend.

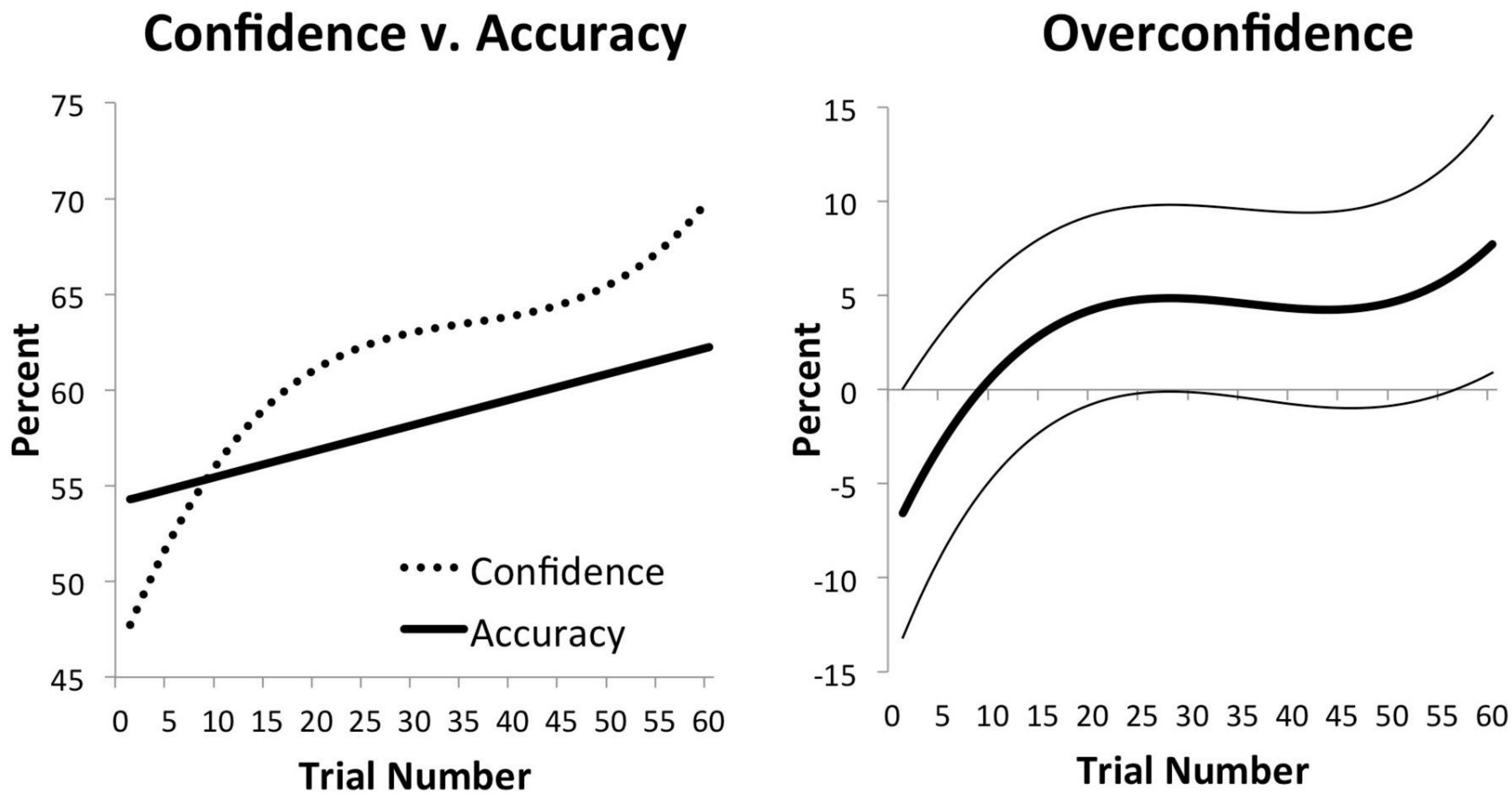


Figure 4. Confidence and accuracy trends over 60 diagnosis trials (Study 4). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

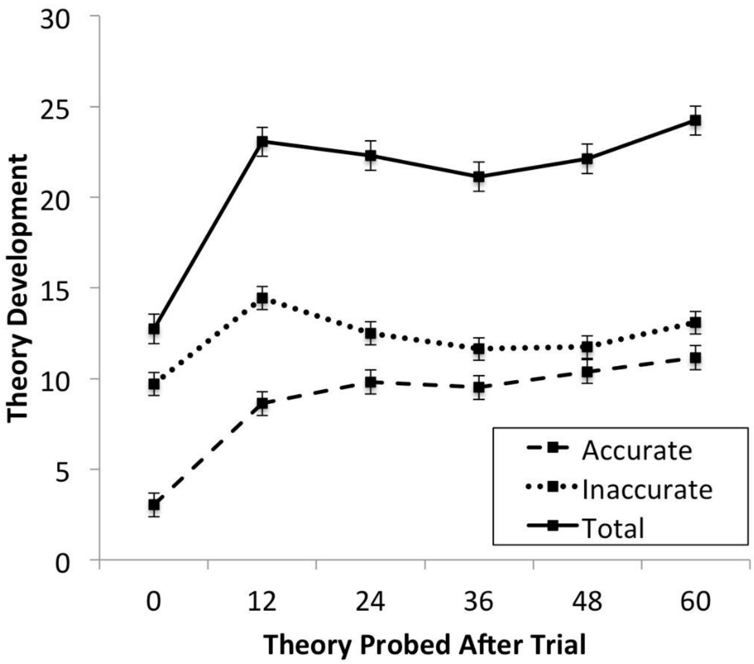


Figure 5. Theory development over experience. The figure displays theory development that is accurate, inaccurate, as well as the sum of the two.

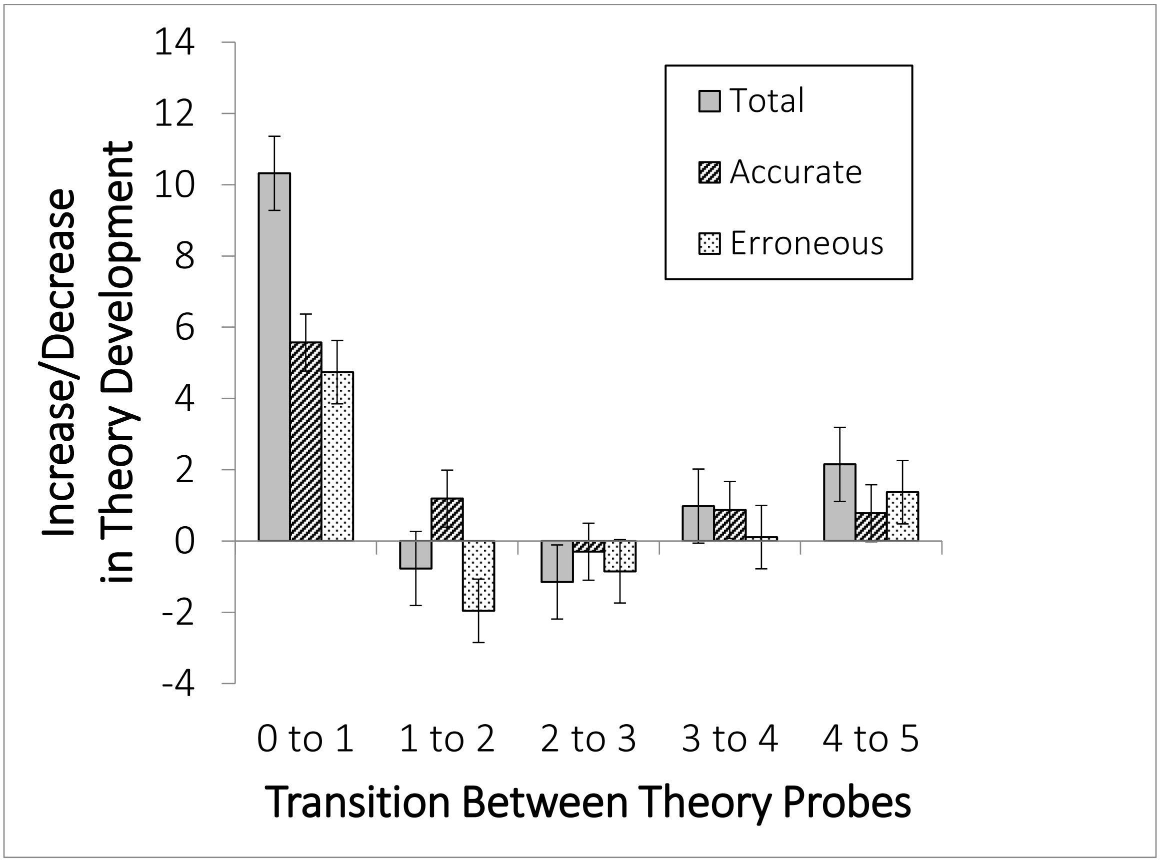


Figure 6. Degree of change in accurate, erroneous, and total theorizing taking place between theory probes.

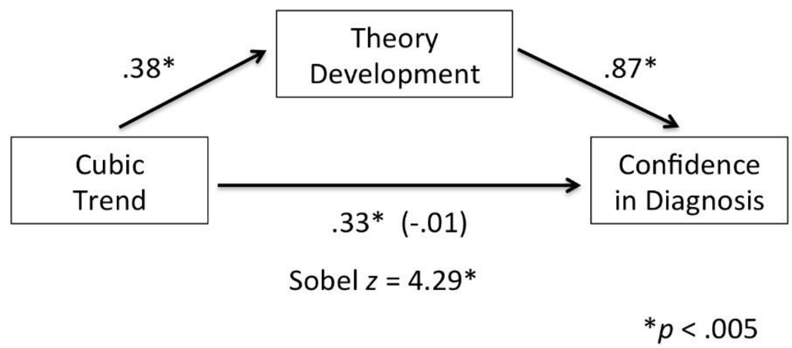


Figure 7. Mediational analysis testing whether the cubic trend in experience in theory development accounts for the cubic trend in confidence.

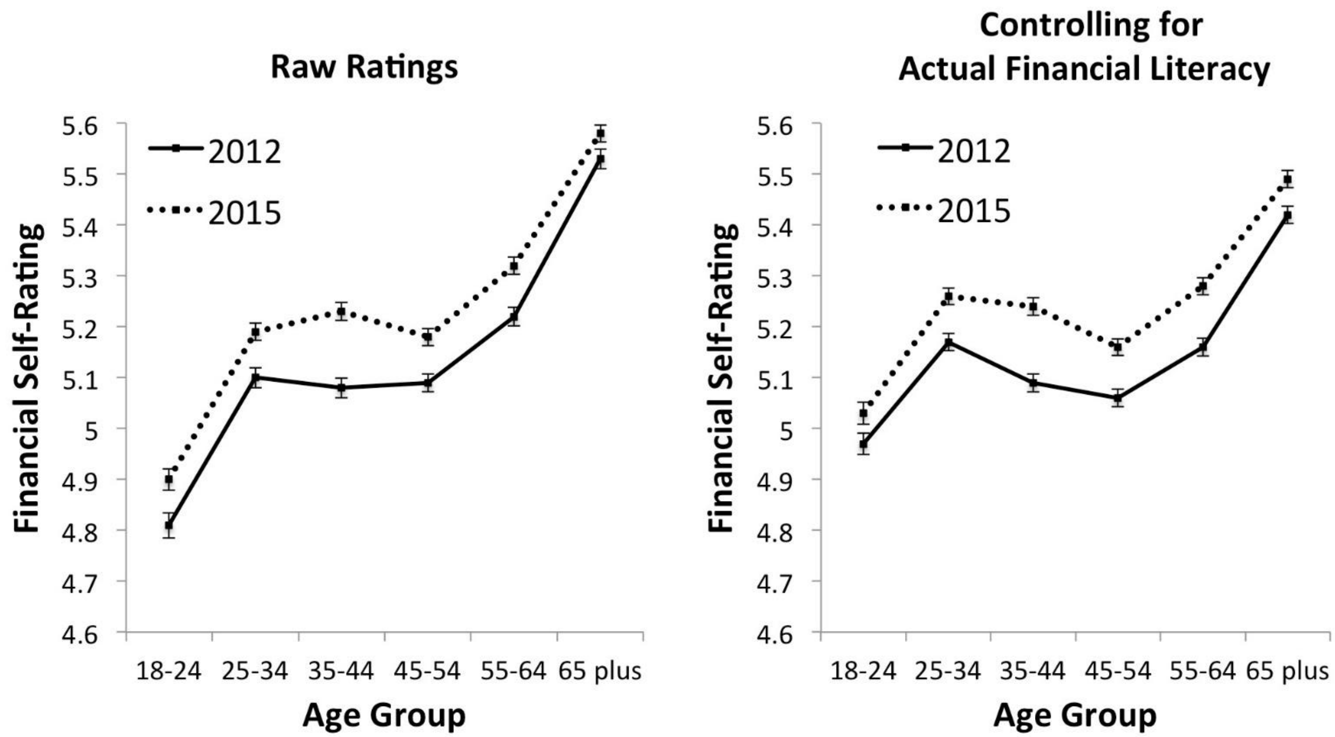


Figure 8. Self-perceived financial literacy (Studies 5a and 5b). . Left Panel: Raw self-ratings. Right Panel: Self-ratings after controlling for actual financial literacy.

Chapter 3: Decision Fluency and Overconfidence among Beginners

Daily life requires people to learn in environments where outcomes are inherently uncertain. Take probabilistic learning, which requires that someone develop an optimal strategy for success, such as in stock picking, playing poker, or finding love, knowing that the outcomes are in some measure due to chance. Probabilistic learning is especially difficult, and yet it characterizes many of the complex challenges people face (Brunswik, 1943; Estes, 1976; Little & Lewandowsky, 2012). Thus, it is not a surprise that people continually confront profound challenges in probabilistic learning, despite broad and deep experience with such tasks.

One challenge is accurately assessing one's chances for success, that is, applying appropriate confidence in one's decisions. People tend to be overconfident (Sanchez & Dunning, 2018, 2019; Sieck & Yates, 2001), and nowhere is this more evident than of beginners. In one recent study, participants played a novel game diagnosing whether patients suffered from a zombie disease (Sanchez & Dunning, 2018, 2020). Their actual learning rate was slow, incremental, and linear. Participants were initially well-calibrated about their accuracy, but they soon became overconfident after only a few trials of the task. That is, subjective learning displayed a "beginner's bubble" of overconfidence that then moderated before resuming an upward climb.

In this manuscript, we explore one account for why beginners start with appropriate confidence but then accelerate rapidly into overconfidence. Specifically, we focused on the relationship between decision fluency and the beginner's bubble of overconfidence, asking whether naturally-occurring changes in fluency predicted this pattern in confidence. Much psychological work shows that people are confident to the extent their decisions are fluent, that is, reached quickly and easily. Such fluency can come from the familiarity with the task stimulus (Fazio, Brashier, Payne, & Marsh, 2015; Kelley & Lindsay, 1993; Koch & Zerback, 2013; Schwartz & Metcalfe, 1992; Unkelbach, 2007), procedural repetition (Pew, 1969; Williams, Duke, & Dunning, 2018), or—as we focused on here—on decision speed (Johnson, 1939; Vickers & Packer, 1982; Audley, 1960; Baranski & Petrusic, 1998; Henmon, 1911; Johnson, 1939; Kiani, Corthell, & Shadlen, 2014; Volkmann, 1934).

As such, we examined whether the beginner's bubble followed a similar pattern in decision speed. It seems reasonable that beginners, new to a task, begin slowly—but then speed up as they become familiar with the task and its procedure, as well as in the strategies they adopt for success. This naturally-occurring speed-up may be connected to confidence, but not necessarily to accuracy. To be sure, decision speed is related to accuracy in some judgments (Festinger, 1943), such as eyewitness identification (Dunning & Peretta, 2002; Robinson, Johnson, & Herndon, 1997; Smith, Lindsay, & Pryke, 2000; Sporer, 1993), but speed does not necessarily connote accuracy in other tasks (Schouten & Bekker, 1967; Wickelgren, 1977; Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). As such, decision speed may possess some constrained ecological validity as an indicator of accuracy, but people might also give it too much weight in assessments of that accuracy.

In total, we predicted that decision speed would follow a curvilinear trend as people gained experience at a task. First, people would rapidly increase their speed at a task as they initially learned it, but this would be followed by a period in which it remained stable. This moderation would arise because people would still be revising and adjusting their approach to the task (see Sanchez & Dunning, 2018), throwing away mistaken ideas while acquiring new ones to improve performance. In short, people would still be in a "learning phase." However, as the task went on, people would gradually slip out of a learning phase into a “performing phase,” in which they were just interested in executing the task as accurately and as efficiently as possible (Denrell, 2005; Einghorn & Hogarth, 1978; Elwin, 2013). Their decisions would once again speed up, and the amount of time they dwelled on feedback, especially feedback after reaching wrong decisions, would also evaporate. As a consequence, their confidence would again rise.

Ultimately, this logic suggests that decision speed would display a cubic trend over time, with two inflection points: a rapid rise at the beginning as participants learned the task, leading to a plateau of speed that would turn into a second rise as participants switched from primarily learning the task to merely performing it. As such, decision speed would match the cubic trend in confidence over time already observed among beginners (Sanchez & Dunning, 2018, 20120). Decision speed, however, would not serve as good an indicator of accuracy in judgment.

Across two studies we tested the relationship between learning, decision times, and confidence by having participants engage in two different probabilistic learning tasks with identical materials and procedures that have been used in previous research (Sanchez & Dunning, 2018, 2020). We predicted a cubic trend over time for decision speed, with that trend mimicking confidence and overconfidence among beginners. In a third study, we focused on intuitions people have about decision speed and confidence. Would their beliefs about the correlation between the two mirror the actual correlation observed?

Method

Participants. For the first two studies participants were recruited from MTurk (both *n* = 100), paid $3, and won an additional $3 if they achieved an overall accuracy level of 80% on the prediction tasks they faced. Written informed consent was obtained from all participants before they began the experiment. We anticipated moderate effects (d = .5) given pilot data, and expected that at a sample size of 52, we had a 95% chance of capturing a significant within-subject finding. However, we rounded up our sample size to 100 given the number of predictions we examined, as well as anticipating the removal of participants who failed manipulation checks.

Procedure and Materials. All studies involving multi-cue probabilistic learning tasks. While they were engaging in these tasks, participants reported their confidence for each diagnosis by assessing the chance from 33% (chance level) to 100% (certainty) that their judgment was correct. In Study 1, participants were asked to imagine they are medical residents in a post-apocalyptic world that has been overrun by zombies. For each of 60 trials, they diagnosed whether a patient had one of two zombie diseases (TS-19 or MZD) or neither, by viewing data about the presence or absence of eight different symptoms. They received feedback on their accuracy after each diagnosis. Refer to the Appendix A for a sample trial.

In Study 2, in a different task, participants imagined they had just invented two lie detection devices that assessed lying differently. They needed to learn which machine handled which set of criteria better. For 60 trials, they were shown criteria which might indicate lying and had to judge which lie detection device (or neither) would work best to tell who was lying. Again, they were given feedback on their judgment after each trial. Thus, the study had a similar overall design as Study 1, but differed in look and in the probabilities connected to outcomes. For additional information on methods and procedure see Sanchez & Dunning, (2018).

Last, in Study 3, we probed participants' intuitions about the relationship between decision speed and confidence. We first familiarized participants with the zombie diagnosis task used in Study 1. Specifically, participants completed 15 trials of the procedure in Study 1. To gauge the relationship participants perceived between reaction times and confidence, we asked them to imagine they would be continuing in this task. They then reported how *confident* they thought they would be if they took 4, 6, 12, and 20 seconds to *diagnose* a patient. They separately answered the same question if they took 1, 1.5, 2, 5, and 14 seconds to look at *feedback* (see Appendix B for exact procedure). Question order (i.e., did participants consider slow or fast times first) was counterbalanced.

Results and Discussion

Participants who never varied their confidence estimates or failed a manipulation check about their understanding of the task were excluded from analyses in Study 1 (n = 5), Study 2 (n = 14) and Study 3 (n = 11).

We pursued four different aims. First, in Studies 1 and 2, we replicated past work (Sanchez & Dunning, 2018, 2020) showing that accuracy would improve in a slow and incremental linear fashion, but that confidence would show a more complex cubic trend starting with a burst, then flattening out, and then beginning to rise again late in the experimental session. Second, we focused on whether decision speed would mimic confidence's cubic pattern, but show less of a relationship to accuracy. Next, we asked whether decision speed statistically mediated the impact of experience on confidence? That is, was the beginner's bubble related to the rapid in increase in decision speed among the initial stages of the learning task? Finally, in Study 3, we assessed how reaction times relate to perceptions of confidence and performance.

Accuracy. To ensure that participants learned with experience and feedback, we conducted a logistic mixed model analysis (random-intercept, random-slope) assigning experience (i.e., trial number) as a fixed variable and participant as a random variable. As seen in Figure 1, accuracy rose in a linear fashion: b = .004, se_b = .002, p = .009, OR= 1.004, BIC = 7656 (Study 1); b = .008, se_b = .002, p < .001, OR = 1.008, BIC=7082 (Study 2). The quadratic term was significant but did not result in a better fitting model in Study 1, b = -.0002, se_b = .0001, p = .05, BIC = 7680, and Study 2, b = -.0004, se_b = .0001, p = .001, BIC = 7103.

Confidence and Overconfidence. Overall, participants proved to be overconfident. Study 1 produced a mean confidence of 64.4% against an accuracy rate of 60.6%, $t(94) = 2.36$, p = .02 In Study 2, overall confidence (M = 64.6) exceeded accuracy (M = 53.9), $t(85) = 6.08$, p < .001 (see Table 1 and Figures 2 and 3). To test more specifically how confidence changed with experience, we conducted a mixed-model regression analysis with linear, quadratic, and cubic terms for experience as fixed effects and participants (random intercept, random slopes) as a random variable. In both studies, all terms were significant (see Table 1), with the cubic model showing the best fit, as measured by BIC.

Decision Speed. Next, we examined how decision speed changed with experience, we looked at time (1) spent making a decision and (2) devoted to examining feedback after reaching that decision. In Study 1, reaction times were strongly skewed (skew = 17.9 and 12.7 for decision and feedback time, respectively) and showed too much kurtosis (kurtosis = 617 and 332 for decision and feedback time, respectively). A reciprocal transformation corrected skew and kurtosis most effectively (skew = .58 and .50 for decision and feedback time, respectively) and kurtosis (kurtosis = .35 and .19, for decision and feedback time, respectively). We then multiplied each transform by 100 to create a decision speed variable.

As expected, decision speed in Study 1 was more closely related to confidence, b = .97, se = .08, p < .001, r_{partial} = .81, than to accuracy, b = 0.002, se = .001, p = .090, r_{partial} = .19, Steiger's z = 10.20, p < .001. In Study 2, decision speed again predicted confidence, b = .74, se =

.09, $p < .001$, $r_{\text{partial}} = .68$, and accuracy, $b = .003$, $se = .001$, $p = .03$, $r_{\text{partial}} = .19$, both in a positive direction. However, the correlation of decision speed to confidence again exceeded that between speed and accuracy, Steiger's $z = 4.82$, $p < .001$.

To test how decision times changed across the task, we conducted a mixed model of analyses (random intercepts and slopes), with time as the dependent variable, examining whether there were linear, quadratic, and cubic terms of time for experience. As predicted, the best fitting relationship for total time spent involved a cubic trend in Study 1, $b = .00009$, $se_b = .00001$, $p < .001$, and Study 2, $b = .00007$, seb $= .00002$, $p < .001$, and the other time variables followed a similar pattern (see Table 2). For decision time, participants started slow but quickly speeded up within a few trials. After this initial phase the rate of increase for decision speed stabilized. Then, near the middle of the task, decision speed began to pick up again. Figure 4 depicts this trend in terms of time spent making decisions.

Time spent with feedback showed the same pattern (again see Figure 5). To examine this variable, we added to the analysis whether participants were right or wrong on the trial, as well as the interaction of being right or wrong with linear, quadratic, and cubic trends over time. Time spent with feedback also followed a cubic trend through time, with participants looking at feedback more overall when they were wrong versus when they were right, $b = 8.32$, $se = 0.48$, $p < .001$, $b = 5.25$, $se = 1.01$, $p < .001$, in Studies 1 and 2, respectively. In Study 1, this main effect also interacted with the quadratic term of time, $b = -.006$, $se = .002$, $p = .002$, though this interaction failed to replicate in Study 2.

We should note that, after roughly 40 to 45 trials, time spent both making decisions and examining feedback both began shortening, suggesting that participants devoted less effort to learning or revision and instead had turned to rote performance. Furthermore, confidence changes at around the same time as decision speed does, indicating that the increasing speed found at the end of the experiment was accompanied by enhanced confidence (see Figure 3).

Mediation. Our last set of analyses explored whether these patterns in decision times related to the overall cubic trend seen on confidence here and before (Sanchez & Dunning, 2018a, 2018b). We have already demonstrated several conditions necessary for mediation: First, a cubic trend predicts both confidence and decision speed in both studies.

To complete a mediational analysis, we examined whether decision speed remained correlated with confidence after the cubic trend was controlled for. Thus, in Study 1, we repeated our multi-level analysis on confidence with decision speed as a covariate. Decision speed continued to be significantly related to confidence, $b = .44$, $se_b = .04$, $p < .001$, though the cubic trend in confidence remained significant even though it was weakened, $b = .0003$, $se_b = .00005$, $p < .001$, consistent with partial mediation, Sobel $z = 6.23$, $p < .001$. A similar analysis examining time spent on feedback failed to establish mediation of the experience/confidence relationship, Sobel $z = .95$, $p = .34$. In Study 2, decision speed continued to predict confidence in a similar analysis, $b = .26$, $se = .03$, $p < .001$, with the cubic trend on confidence being reduced, $b = .0002$, $se = .00008$, $p = .01$, Sobel $z = 3.16$, $p = .002$, but not eliminated. Such a pattern is consistent with partial mediation (see Figure 4).[9, 10]

[9] We should caution the reader that we cannot draw causal inferences, as decision speed was not directly manipulated. We suggest that decision speed leads to changes in confidence because it occurs prior to the confidence estimate in each trial. However, further tests of mediation are consistent with the reverse notion that confidence mediates the relation of time trends in decision speed. In study 1, Sobel $z = 4.41$, $p <.001$, and study 2, Sobel $z = 2.69$, $p =.007$. These results are also consistent with the assertion that speed and confidence are both produced in tandem from some common cause. To be sure, neither of these interpretations preclude our assertation that decision speed has a greater relationship with confidence than it serves as a valid predictor of accuracy.

[10] An observant reviewer questioned whether we would still obtain the same pattern of results if participants did not know when the study would end. It could be that the second inflection point in confidence occurs near the end of the study because participants speed up when they know the study will end. As such, we conducted a replication of Study 1 (n=100, 11 participants failed a manipulation check or never varied in their confidence indicating they ignored the measure). This study did not contain a "progress bar" and participants were not told how many patients they would diagnose. Overall we replicated our effects. Confidence continued to follow a cubic function, $b =$

Last, for exploratory reasons, in Study 3 we tested whether participants correctly perceived that faster decision times are related to greater confidence. We found that participants believed that if they took longer to diagnose a patient, $r = .44$, $p < .001$, and examining feedback, $r = .41$, $p < .001$, it would indicate they would be more confident in their diagnosis. Further, they reported that taking longer to diagnose a patient r = .42, p < .001, and to examine feedback, r = .34, p < .001, would suggest that they were more likely to reach a correct diagnosis. In sum, participants perceived slower reaction times to serve as a positive indicator of confidence and accuracy, the exact opposite of what we found when participants made real diagnoses.

Summary. In sum, we found that as people learn, the time spent on each trial rapidly decreases, eventually levels off, and then begins to decrease again as people gain even more extensive experience with a task. This pattern of reaction times predicted patterns in confidence as people gain experience at a novel task. Specifically, how long participants reviewed feedback did not predict this pattern of confidence.

General Discussion

The goal of the present research was to examine the decision-making process of beginners, starting with the question of why confidence and accuracy so quickly dissociate as people gain experience with a novel multi-cue probabilistic learning task. In two studies, we replicated a cubic trend in judgmental confidence as people tackled a new task, displaying by a "beginner's bubble" of rapidly rising overconfidence after only a few trials of experience. Across our studies we also demonstrated a similar naturally-occurring curvilinear pattern in decision

---

.0004, $se = .00006$, $p < .001$, as well decision speed, $b = .0001$, $se = .00002$, $p < .001$. Also, decision speed, Sobel $z = 3.91$ $p < .001$, mediated the cubic trend in confidence. Unexpectedly, feedback speed also explained the beginner's bubble, Sobel $z = 2.00$ $p = .046$.

speed. As participants gained experience with a learning task, their decision speeds increased markedly, but then leveled off, only to begin rising again after substantial experience with the task, as they apparently moved from a learning phase to a performance one. As we predicted, this cubic trend in decision time, at least in part, anticipated the pattern seen in confidence and overconfidence among our participants. In short, these findings suggest that the connection between experience and decision speed is similar to the pattern seen in confidence, specifically the beginner's bubble in overconfidence (Sanchez & Dunning, 2018, 2020).

Importantly, this research also provided data addressing a different but important issue concerning the wisdom of using reaction time as an indicator of accuracy. Should a person rely on speed as a positive indicator of a judgment's quality? Our data suggest one should be cautious in doing so. The relationship between decision speed and accuracy appears is significant: Across the first two studies, speed was related to accuracy, $z = 2.70$, $p = .006$. However, the relationship of decision speed to confidence far overstated the magnitude of that relationship, $z = 12.99$, $p < .001$. Further cautioning on the use of decision speed to infer accuracy, other researchers have shown that one can cleanly dissociate speed or fluency from decision accuracy via extraneous circumstances (e.g., Kelley & Lindsay, 1993; Schwartz & Metcalfe, 1992; Unkelbach, 2007; Williams et al., 2018), although they have admittedly used artificial laboratory techniques to do so. Here, we have showed how one quite natural influence on decision fluency, namely experience at a task, can naturally lead to inappropriate confidence.

In a final twist, we found that people's intuitions about decision speed and accuracy may generally be in error. Study 3 showed that intuitions about decision speed ran counter to any true relationship to accuracy. Participants in that study forecast that they would express more confidence, not less, if they were slower in their decisions. However, in actual decisions, the actual relationship between accuracy (and expressed) confidence ran in the opposite direction.

That said, we caution that we cannot claim a causal role for decision speed in determining confidence. Although it predicts confidence it may not cause that confidence. To be sure, if there is a causal relationship, it is arguably more plausible that it is decision speed that prompts confidence than the reverse, with confidence predicting decision speed, in that the speed of a decision in each trial occurs in time prior to the confidence judgment.[11] However, our data do not rule out a "common cause" model. That is, there may be some hidden third variable that causes decision speed and confidence to wax and wane in parallel, thus allowing speed to predict confidence. One variant of this possibility is that confidence and accuracy actually represent different facets of the same variable, with speed being a behavioral measure of confidence rather than an input into it. As such, it remains for future research to specify further the relationship among experience, decision speed, and confidence. Such work could directly manipulate reaction times, or confidence, to test which way the causal arrow might go. Alternatively, future research may search for a common cause for the two variables.

In the meantime, the studies presented here make several contributions to the existing literature on overconfidence. First, the studies replicate previous findings about how confidence grows, and far outpaces accuracy rates, among beginners (Sanchez & Dunning, 2018, 2020),

[11] To be sure, one might argue that confidence acquired in prior trials may predict later decision speed, so we conducted a Granger causality analysis based on time series data across trials from Studies 1 and 2 (Feige & Pearce, 1979). We first detrended the data by looking at differences in confidence and decision speed from trial to trial to remove autocorrelation. An initial analysis suggested the best models to predict decision speed from previous decision speed included data from one and two trials prior. We then added confidence from those trials to see if it enhanced the prediction of later decision speed. It did not: Fs(2, 54) = 1.84 and .26, ps > .05, for Study 1 and 2, respectively. Thus, earlier confidence failed a test examining whether it caused later decision speed. A reverse analysis showed that earlier decision speed also failed to predict later confidence, Fs(2, 54) = .16 and 3.13, ps > .05, for Study 1 and 2, respectively. However, adding decision speed from the same trial to the analysis enhanced the prediction of confidence, Fs(1, 53) = 4.70 and .26, ps > 8.26, for Study 1 and 2, respectively. Thus, contemporaneous decision speed once again predicted confidence, after controlling for earlier confidence and decision speed.

with subjective confidence on a two-cue prediction task exhibiting a much more complex change pattern across 60 trials of experience than accuracy. As a consequence, overconfidence bears a much stronger statistical correlation to confidence than to accuracy. Second, decision speed shows pattern that mirrors that of confidence. Finally, people believe that accuracy increases markedly with decision time, but the opposite is the case.

References

Audley, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review*, *67*(1), 1-15.

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*(3), 929-945.

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, *33*, 10-16.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*(3), 255-272.

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112,* 951–978.

Dunning, D., & Perretta, S. F. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from erroneous positive identifications. *Journal of Applied Psychology, 87*, 951-962.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Elwin, E. (2013). Living and learning: Reproducing beliefs in selective experience. *Journal of Behavioral Decision Making, 26,* 327–337.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37–64.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*, 993-1002.

Feige, E. L., & Pearce, D. K. (1979). The casual causal relationship between money and income: Some caveats for time series analysis. *The Review of Economics and Statistics*, *61*, 521-533.

Festinger, L. (1943). Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, *32*, 291-306.

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes*, *98*, 39-53.

Henmon, V.C.A. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, 18, 186–201.

Johnson. D.M., 1939. Confidence and speed in the two-category judgment. *Archives of Psychology*, 34, l-53.

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.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*, 1329-1342.

Koch, T., & Zerback, T. (2013). Helpful or harmful? How frequent repetition affects perceived statement credibility. *Journal of Communication*, *63*, 993-1010.

Little, D. R., & Lewandowsky, S. (2012). Multiple-cue probability learning. In *Encyclopedia of the Sciences of Learning* (pp. 2386-2388). Springer US.

Pew, R. W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica*, *30*, 16-26.

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, *82*, 416-425.

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114,* 10-28.

Sanchez, C**.** & Dunning, D. (2019). Cultural patterns in student assessments and performance explain the worldwide perception/performance paradox in math and science skill, *Social Psychological and Personality Science*, *10,* 935-945.

Sanchez, C., & Dunning, D. (2020). *The jumping-to conclusions: Implications for reasoning, errors, false beliefs, knowledge corruption, and impeded learning*. Manuscript submitted for publication.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta psychologica*, *27*, 143-153.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1074-1083.

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1003-1021.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed?. *Journal of Applied Psychology*, *85*, 542-550.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*, 22-33.

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179-197.

Volkmann, J. (1934). The relation of time of judgment to certainty of judgment. *Psychological Bulletin*, 31, 672–673.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, *41*, 67-85.

Williams, E. F., Duke, K. E., & Dunning, D. (in press). Consistency just feels right: Procedural fluency increases confidence in performance. *Journal of Experimental Psychology: General.*

Appendix A: Sample Trial from Study 1

Throughout this paper multicue probabilistic learning tasks were used. For illustrative purposes, we include a sample trial depicted from Study 1. First participants were presented with a diagnosis. Then they received immediate feedback,

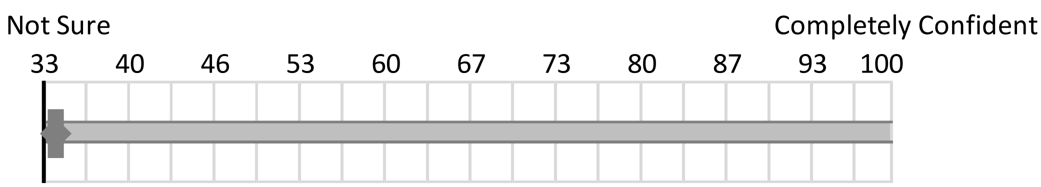
Diagnosis:

Symptoms	**Present?**
Congestion	no
Brain Inflammation	no
Abscess	yes
Swollen glands	yes
Itching	no
Rash	no
Fever	no
Glossy eyes	no

What is the patient's diagnosis?

MZD TS-19 Neither

Please report how confident you are in this decision. What's the chance that you are right, from 33% to 100%? Mark 33% if you think it's just as likely that you are wrong as you are right (I.e., it's 33-33-33 that I'm right). Mark 100% if you are absolutely sure that you are right; there's no chance that you are wrong. Mark 66% if you think the chance that you are right is 2 out of 3. Mark whichever probability best indicates the specific chance that you are right.



Feedback if they were incorrect:

You made an error.

This patient had neither.

Symptoms	**Present?**
Congestion	no
Brain Inflammation	no
Abscess	yes

Swollen glands	yes
Itching	no
Rash	no
Fever	no
Glossy eyes	no

Feedback if they were correct:

Correct.

This patient had neither.

Symptoms	Present?
Congestion	no
Brain Inflammation	no
Abscess	yes
Swollen glands	yes
Itching	no
Rash	no
Fever	no
Glossy eyes	no

Appendix B: Procedure Used in Study 3

Below is the procedure used in Study 3.

Participants were told the following: *We would like you to think about the time it took you to perform this task. Specifically think about the following: 1. How long it took you to reach a decision/diagnose the patient 2. How long you spent looking at the feedback. We have timed people to see how many seconds it takes them to diagnose the zombie patients. Imagine you will be continuing to diagnose patients in this task as you answer the following questions. The first set of questions refers to how quickly it took you to make the medical diagnosis when you saw the medical symptoms. The typical person spends around 8 seconds doing this. Tell us if you were to continue in this task... How confident you would be right now if we showed you another patient and it took you the following amount of seconds to select a diagnosis.* Then participants reported on a slider, labeled *Medical diagnosis time in seconds,* from 33(not sure) to 100 (completely confident) they would be if they spent 4, 6, 8, 12, and 20 seconds.

Then they were instructed the following: *Next we want you to imagine you just diagnosed the patient AND reported how confident you were. We want you to think about how long you spent looking at feedback afterwards. The typical person spends around 2 seconds doing this. Tell us if you were to continue in this task.... How confident you would have been on the PRIOR SCREEN, if we showed you another patient and you spent the following amount of time looking at feedback.* Then participants reported on a slider, labeled *Time spent looking at feedback in* *seconds,* from 33(not sure) to 100 (completely confident) they would be if they spent 1, 1.5, 2, 5, and 14 seconds.

In a similar manner participant answers questions related to their accuracy. They were instructed the following: *Now we want you to think about whether you would be accurate or not. What is the likelihood you would be CORRECT if you spent the following amount of time diagnosing a patient. The typical person spends around 8 seconds doing this.* Then participants reported on a slider, labeled *Medical diagnosis time in seconds,* from 33(correct by chance) to 100 (definitely correct) they would be if they spent 4, 6, 8, 12, and 20 seconds.

Lastly, participants were instructed the following: *Next we want you to think about how long you spent looking at feedback. The typical person spends around 2 seconds doing this. What is the likelihood that you would have been correct IN YOUR PRIOR DIAGNOSIS, if you then spent the following amount of time looking at feedback?* Then participants reported on a slider, labeled *Time spent looking at feedback in seconds,* from 33(correct by chance) to 100 (definitely correct) they would be if they spent 1, 1.5, 2, 5, and 14 seconds.

Table 1.

Linear, Quadratic, and Cubic Model Predicting Confidence from Experience (Studies 1 and 2)

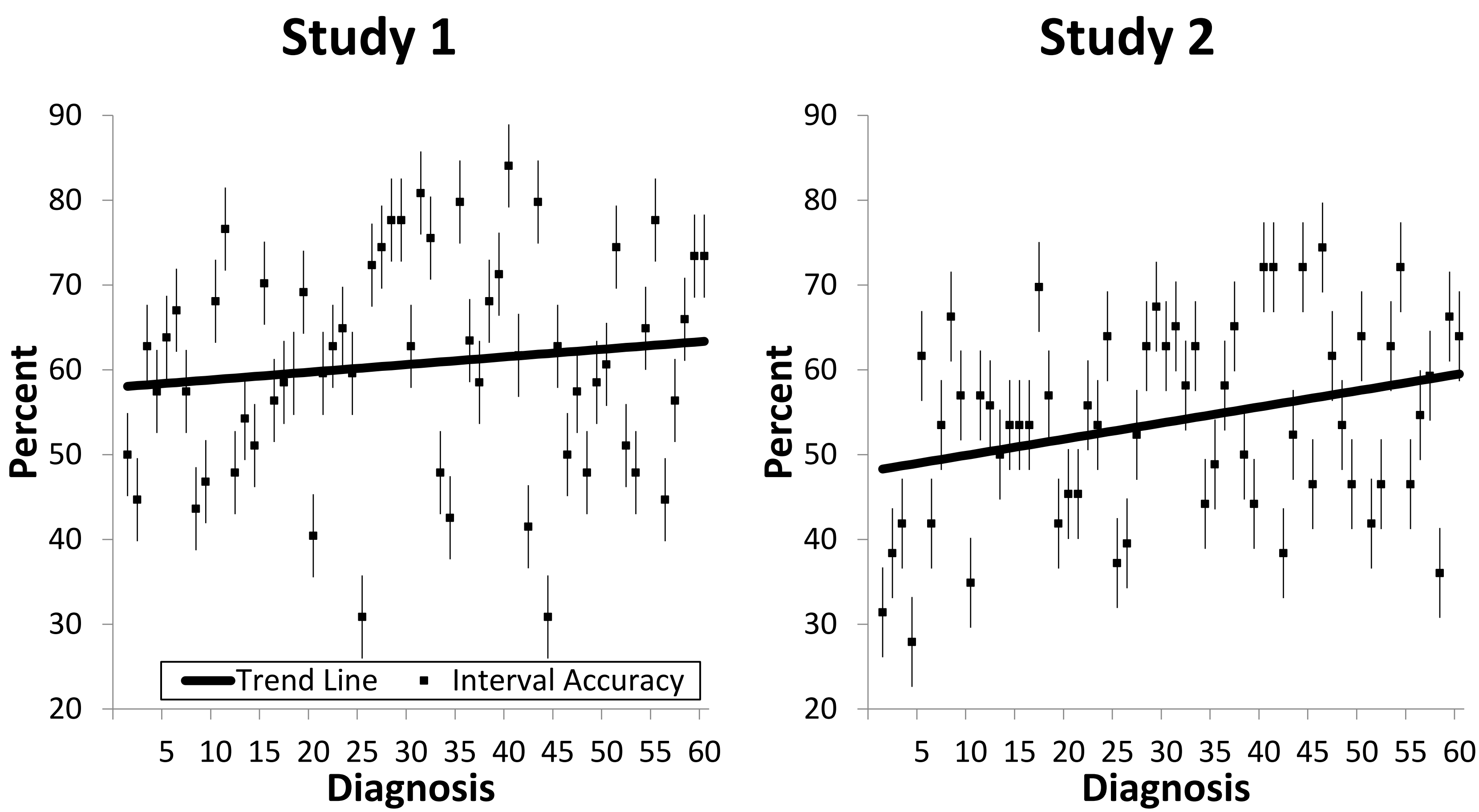
				Measure			
Model	*b*	se_b	*p*	*95% Lower*	*95% Upper*	η_p^2	*BIC*
			Study 1				
Linear	.28	.03	< .001	.22	.34	.70	46802
Quadratic							
Linear	.28	.03	< .001	.22	.34	.70	46928
Quadratic	-.01	.001	< .001	-.01	-.01	.64	
Cubic							46214
Linear	.11	.05	.03	.01	.21	.22	
Quadratic	-.01	.001	< .001	-.01	-.01	.64	
Cubic	.0003	.0001	< .001	.00	.00	.44	
			Study 2				
Linear	.19	.03	< .001	.13	.25	.56	41751
Quadratic							
Linear	.19	.03	< .001	.13	.25	.56	41554
Quadratic	-.006	.001	< .001	-.01	-.003	.43	
Cubic							41459
Linear	.07	.06	.21	-.04	.19	.13	
Quadratic	-.006	.001	< .001	-.001	-.003	.43	
Cubic	.0002	.0001	< .001	.0001	.0004	.29	

Table 2.

Linear, Quadratic, and Cubic Models Predicting Decision and Feedback Speed from Experience (Studies 1 and 2)

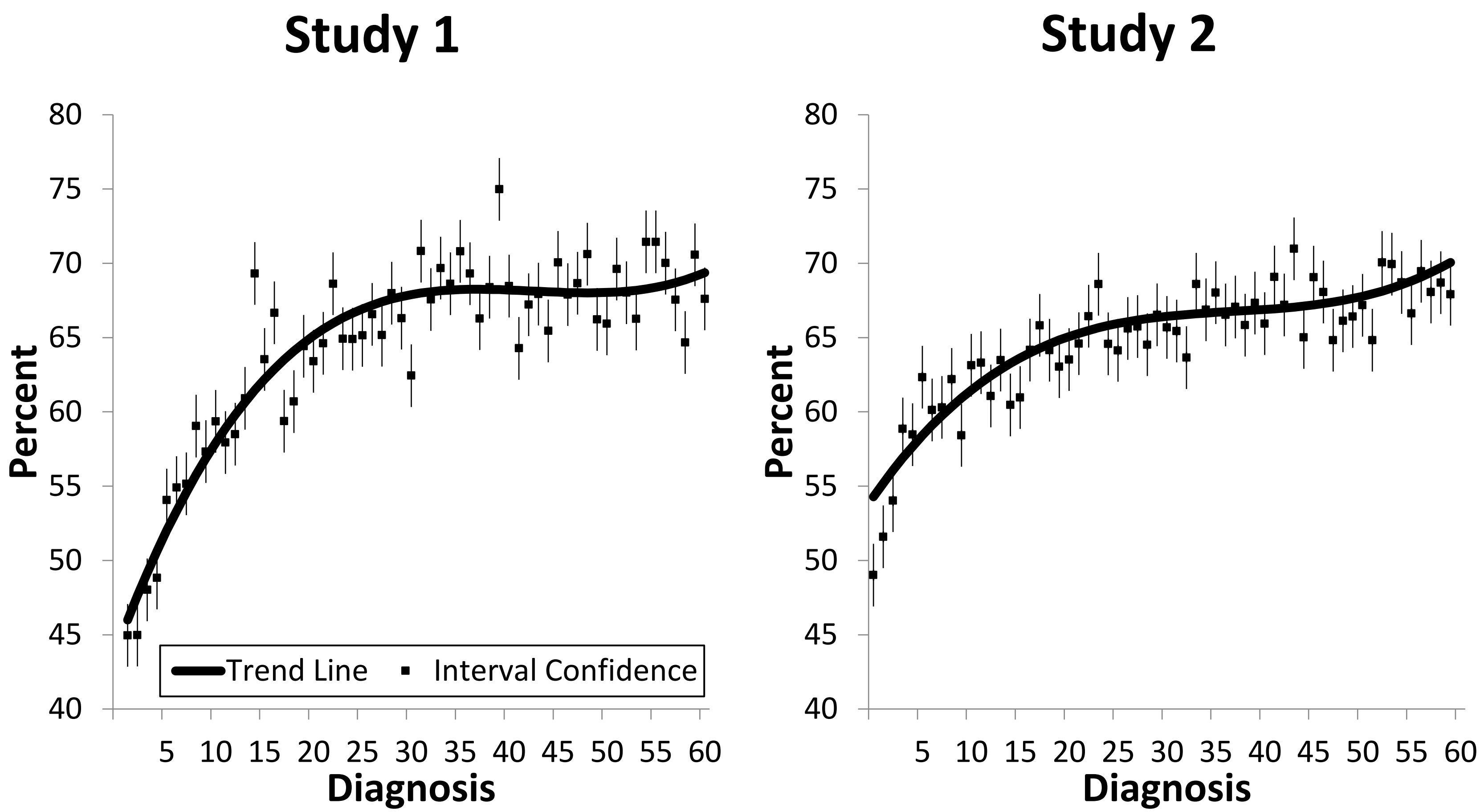
	Measure						
	b	se_b	*p*	*95% Lower*	*95% Upper*	η_p^2	*BIC*
			Study 1				
			Decision Speed				
Linear	.11	.01	< .001	.10	.13	.83	33270
Quadratic							
Linear	.11	.01	< .001	.10	.13	.83	33075
Quadratic	-.002	.0004	< .001	-.003	-.001	.53	
Cubic							
Linear	.05	.01	< .001	.03	.08	.01	33029
Quadratic	-.002	.0004	< .001	-.003	-.001	.0002	
Cubic	.0001	.00002	< .001	.0001	.0001	.00001	
			Feedback Speed				
Linear	0.67	0.04	< .001	0.58	0.76	0.85	50323
Quadratic							
Linear	0.67	0.04	< .001	0.58	0.76	0.85	49801
Quadratic	-0.02	0.002	< .001	-0.02	-0.01	0.67	
Cubic							
Linear	0.438	0.07	< .001	0.29	0.57	0.53	49697
Quadratic	-0.02	0.002	< .001	-0.02	-0.01	0.67	
Cubic	0.0004	0.00009	< .001	0.0003	0.0006	0.44	
			Study 2				
			Decision Speed				
Linear	.12	.01	< .001	.09	.14	.72	33979
Quadratic							33882
Linear	.12	.01	< .001	.09	.14	.72	
Quadratic	-.003	.001	< .001	-.004	-.002	.49	
Cubic							33884
Linear	.07	.02	< .001	.03	.10	.36	
Quadratic	-0.003	0.001	< .001	-0.004	-0.002	.49	
Cubic	.00009	.00003	0.08	0.000	0.000	.34	
			Feedback Speed				
Linear	.65	.08	< .001	.49	.81	.65	53116
Quadratic							53108
Linear	.65	.08	< .001	.49	.81	.65	
Quadratic	-.0133	.002	< .001	-.02	-.009	.58	
Cubic							52923
Linear	.40	.12	< .001	.16	.64	.34	
Quadratic	-.01	.004	< .001	-.02	-.006	.37	
Cubic	.0005	.0002	.03	.00005	.0009	.24	

Figure 1. Time Course of Accuracy



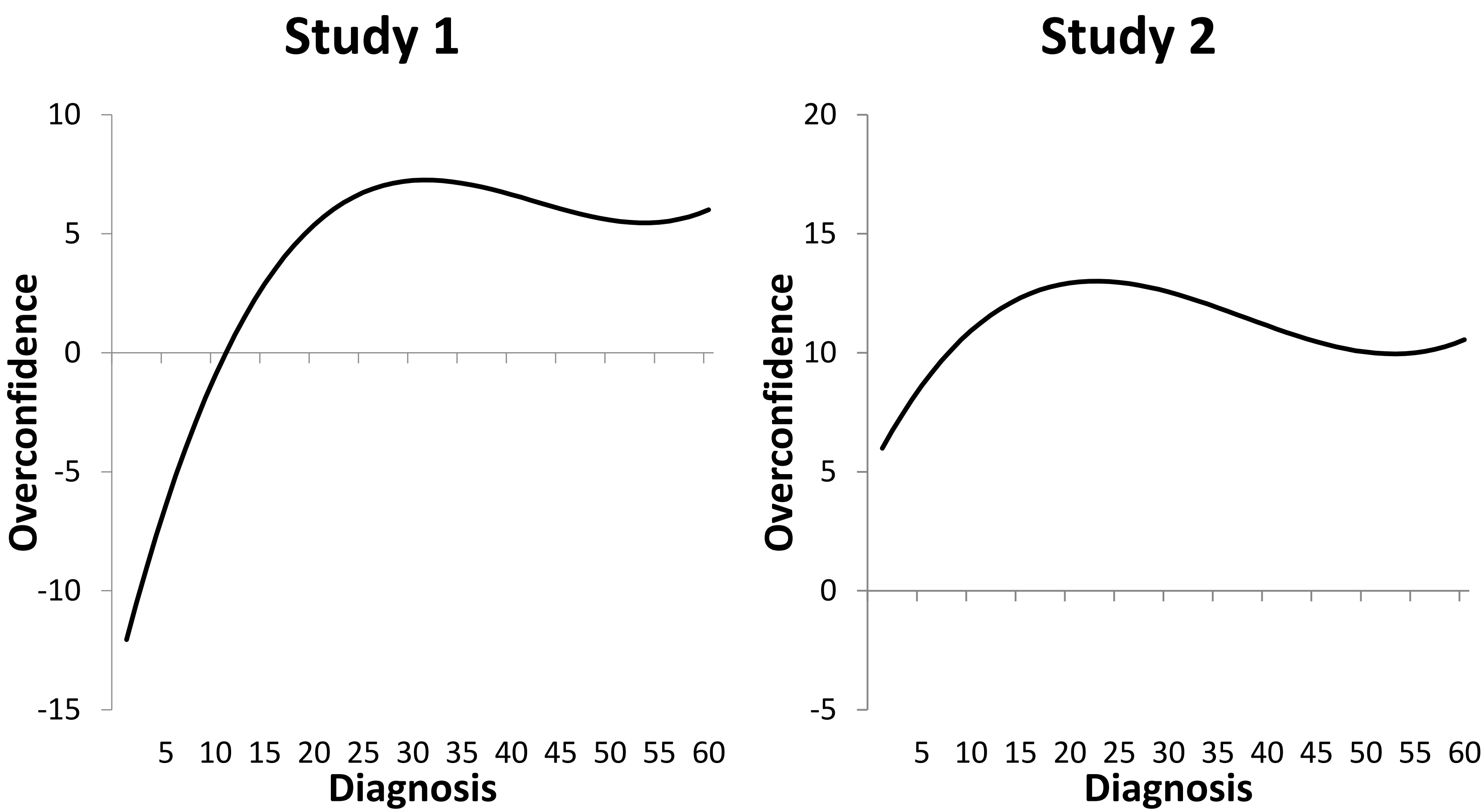
Note: Error bars depict standard error for the mean at each interval.

Figure 2. Time Course of Confidence



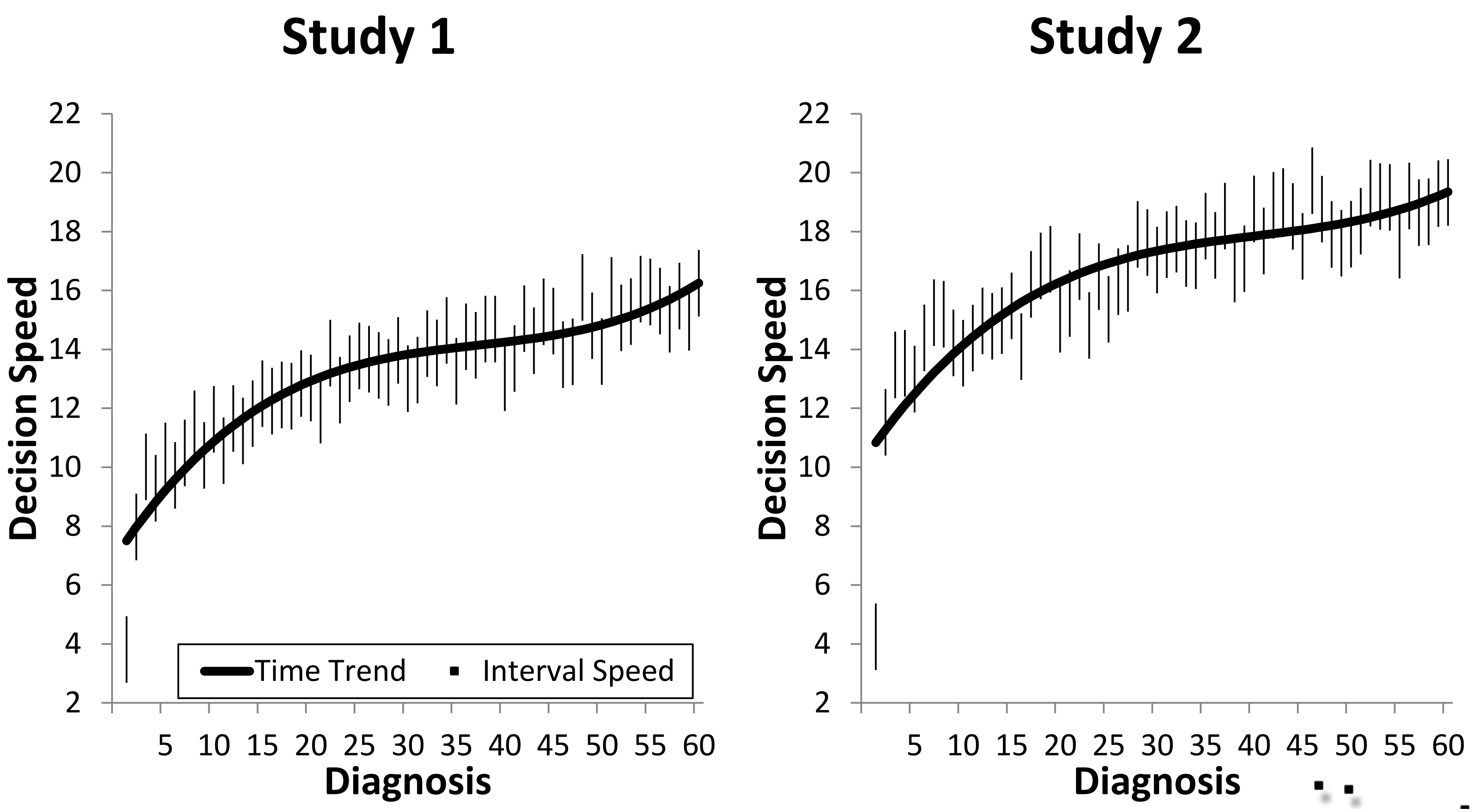
Note: Error bars depict standard error for the mean at each interval.

Figure 3. Time course for differences between fitted values for confidence and accuracy (Overconfidence)



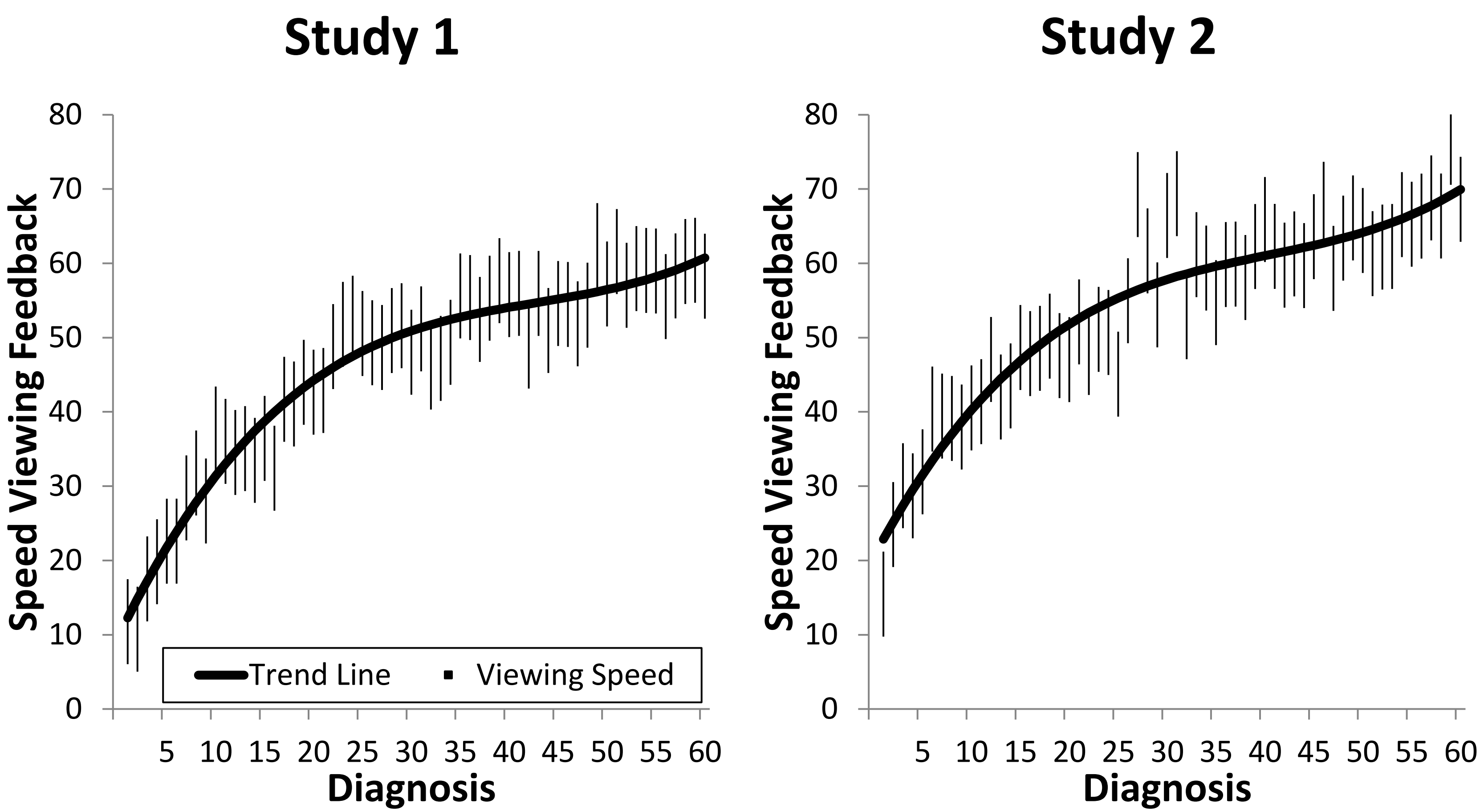
Note: The overconfidence curves represent the fitted values for confidence minus fitted values for accuracy in each study.

Figure 4. Time Course of Decision Speed



Note: Decision speed calculated as (1/reaction time) * 100. Error bars depict standard error for the mean at each interval.

Figure 5. Time Course of Speed Viewing Feedback



Note: Speed calculated as (1/time spent viewing) * 100. Error bars depict standard error for the mean at each interval.

Chapter 4: Jumping-to-Conclusions: Implications for Reasoning Errors, False Belief, Knowledge Corruption, and Learning

Deciding how much information to gather before reaching a decision is a dilemma everyone faces. This dilemma may be especially difficult in tasks that involve risk or doubt because those tasks require people to sort through evidence and reach a conclusion about how to perform a task that is inherently uncertain. Given the difficulty in making such decisions, it is no surprise that ample research suggests that people commit a host of systematic errors when making judgments under uncertainty, with people often unduly confident in their ability to navigate such complex tasks (Dunning, Griffin, Milojkovic, & Ross, 1990; Griffin & Tversky, 1992; Lichtenstein, Fischhoff, & Phillips, 1977).

In this manuscript, using inspiration from decades of research in clinical psychology, we adopt an individual difference approach to gauge how those drawn from a nonclinical population might be those most vulnerable to the types of errors, illusions, impairments, and overconfidence that characterizes everyday judgment and decision-making. In particular, we focus on a specific behavior identified in schizophrenia research that predicts those who are more susceptible to misjudgments and unwarranted confidence.

We begin by noting an important theme from social psychological research related to errors in judgment and decision making: People are inattentive to the reliability of the evidence they have in hand and exaggerate the degree to which a small sample of evidence is representative of ground truth (Benjamin, Rabin, & Raymond, 2016; Griffin & Tversky, 1992; Tversky & Kahneman, 1971; Williams, Lombrozo, & Rehder, 2013). Often, people place far too much emphasis on the first piece of information they see and give this scant evidence undue weight on subsequent theorizing (Asch, 1946; Jones, Rock, Shaver, Goethals, & Ward, 1968;

Kelley, 1950), obfuscating true patterns that might arise in the world (Kamin, 1968; Yarritu, Matute, & Luque, 2015).

In fact, decades of research have shown that people are very comfortable taking small and impoverished amounts of data to reach confident theories about their worlds (Heider & Simmel, 1944; Risen, Gilovich, & Dunning, 2007; Sanchez & Dunning, 2018). They infer too much from short sequences and read meaningful patterns into random or chaotic data (Camerer, 1987; Chapman & Chapman, 1969; Guthrie, Weber, & Kimmerly1993; Ono, 1987; Rabin, 2002). In short, when people develop theories they do so based on the few pieces of evidence they see early on in a task, failing to temper their theorizing given how little information they have (Griffin & Tversky, 1992).

Jumping-to-Conclusions

With this theme in mind, it is worthwhile to turn to research on schizophrenia. Biases and cognitive distortions have long been studied in schizophrenia patients and an exaggeration of them has been found as a pathogenesis of schizophrenia (Bell, Halligan, & Ellis, 2006; Bentall, 1992; Garety & Freeman, 2013; Moritz, et al., 2014; Savulich, Shergill, & Yiend, 2012; Van Der Gaag, 2006). Schizophrenia is a chronic, severe, and rare mental condition, affecting roughly 1% of the population (Cooper et al., 2007). People with schizophrenia experience disabling symptoms that affect how they think, feel, and behave. Amongst these symptoms are delusions and hallucinations (American Psychiatric Association, 2013).

Deficits in probabilistic reasoning have been hypothesized to produce symptoms in schizophrenia patients. Unfortunately, the relationship is not fully understood because schizophrenia is a relatively rare condition and it takes considerable resources to establish causal clarity in the biological, psychological, and social mechanisms in play.

Delusions, which are often characterized as firmly held beliefs with little evidence to substantiate these beliefs, are a prominent and debilitating feature of schizophrenia (Balzan, 2016 Moritz et al., 2014). They are believed to occur, in part, because of deficits in probabilistic learning. For example, delusions can occur when people derive patterns from two unrelated events that happen to occur together (Moritz et al., 2014). In addition to forming beliefs in this manner, schizophrenia patients are also often overconfident in these types of judgments they formulate (Balzan, 2016; Köther, et al., 2012; Moritz & Woodward, 2006a). Provided that people are more likely to act on overconfident judgments (Barber & Odean, 2001; Moritz & Van Quaquebeke, 2014), understanding how overconfidence develops is of importance to schizophrenia researchers.

That said, considerable research over the past two decades has extensively documented one link between cognitive habits and impaired judgment and delusional belief. That tendency is known as jumping-to-conclusions behavior among schizophrenia patients (Dudley, Taylor, Wickham, & Hutton,2015; Garety & Freeman, 1999; Huq et al., 1988; Moritz & Woodward, 2005). Jumping to conclusions behavior refers to whether schizophrenia patients collect only a few pieces of evidence before reaching a decision (Dudley et al., 2015; Moritz & Woodward, 2005). In a common task, participants are shown two urns that contain beads of two different colors, such as red and green, but in different proportions. One urn may have 70 green and 30 red beads and the other may contain the opposite proportion (30 green and 70 red). The experimenters choose one urn and a bead is drawn from the urn, with the patient’s task to decide which urn the experimenter is drawing from, the one with majority-red or majority-green. The experimenter continues to draw beads from that urn, one at a time, until the participant says to stop and states which urn the beads are being drawn from.

In contrast to nonclinical controls, schizophrenia patients, particularly delusion-prone ones, gather very little information before they reach a conclusion in the urn task, often rendering a decision after only one or two balls have been drawn (Dudley et al., 2015; Evans, Averbeck, & Furl, 2015; Freeman, Pugh, & Garety, 2008). These patients are said to be jumping to conclusions (JTC). To be sure, JTC behavior is commonplace, not inherently pathological, and not all schizophrenia patients demonstrate this tendency. There is nothing about schizophrenia that necessarily produces the behavior, and nonschizophrenia patients can show it. That said, roughly half of schizophrenia patients jump to conclusions in decision making tasks (i.e., state a decision after only one or two draws from the urn), compared with 10-20% of other clinical or nonclinical respondents (Fine et al., 2007; Freeman, 2007; Freeman, Pugh, & Garety, 2008). Further, this behavior has been shown to be temporally stable characteristic of the individual. Test-retest reliability has been shown to be .88 for a three-month and .84 for a twelve-month period (So et al., 2012).

Goals of the Present Research

In the research contained herein, we imported research on JTC behavior into the study of a nonclinical population, asking what implications it might have for judgment error, overconfidence, and learning. As such, we did an empirical survey of the correlates of JTC behavior. Would it predict faulty reasoning, judgmental errors, and misplaced confidence in a nonclinical population? We addressed this general question by pursuing several specific aims.

Judgment Errors

First, we proposed that JTC behavior might amplify a person's vulnerability to decision contaminants that occur early in the decision process, like automatic phenomena that are inspired by "system 1" reasoning. This type of thought that is quick, associationistic, intuitive, and relatively effortless (Kahneman, 2011). Researchers have suggested that system 1 is responsible for such biases as belief bias in logical reasoning (Evans, Barston, & Pollard, 1983) and denominator neglect in risky choice (Reyna & Brainerd, 2008). We wished to see if people who exhibited more JTC behavior in a probabilistic reasoning task were more vulnerable to these types of biases among a nonclinical population.

Unrealistic Beliefs

Second, we examined whether the JTC behavior is associated with endorsing impossible, unreasonable, or absurd beliefs like delusions (e.g., "electrical devices can influence the way I think"; "I am being persecuted"), conspiracy theories, and belief in the paranormal. In this, we aimed to replicate and extend previous forays that researchers have conducted exploring the implications of JTC behavior for the general public. Specifically, several studies have already connected JTC with delusional ideation, paranoia, and oddball beliefs in the general population (Freeman, Pugh, & Garety, 2008; Ross, McKay, Coltheart, & Langdon, 2015; Warman, Lysaker, Martin, Davis, & Haudenschield, 2007; Zawadski et al., 2012, although see Ross, Pennycook, McKay, Gervais, & Landon, 2016). We re-examined this question, and aimed to go beyond past work to provide some theoretical understanding for any connection to oddball beliefs we found.

Knowledge Corruption

We also examined whether JTC behavior is related to "knowledge corruption," that is, errors and biases in judgment that people fail to sense or anticipate in themselves. Schizophrenia patients are less calibrated than non-patients about their abilities and weaknesses (Gaweda et al., 2015; Moritz & Woodward, 2002; Moritz, Peters et al.,1999). In recall tasks, they are less confident in their correct memories yet endorse memory errors with more overconfidence (Moritz & Woodward, 2002; Moritz & Woodward, 2006a; Moritz & Woodward, 2006c). It has been suggested that these confidence gaps result in schizophrenia patients facing greater difficulty in distinguishing between correct and incorrect subjective facts, which can promote delusional thoughts (Moritz& Andreaou, et al., 2014; Moritz & Ramdini, et al., 2014; Moritz & Woodward, 2006b).

In the judgment and decision-making literature, these patterns of knowledge corruption would be referred to as overconfidence and miscalibration (Dunning et al., 1990; Lichtenstein et al., 1977). Thus, we asked participants to answer general knowledge questions and to estimate the likelihood that their answers were right. We anticipated that JTC behavior would be linked to exacerbated overconfidence and greater miscalibration about which specific answers were right versus wrong.

Impeded Learning

Next, in two studies, we examined performance and overconfidence in a learning context in which participants learned how to make judgments about uncertain events (e.g., medical diagnoses). Recent research has shown that beginners at a novel task initially express appropriate and cautious confidence about their performance, but rapidly become overconfident as they gain just a little experience at the task—a phenomenon labeled the “beginner’s bubble” in overconfidence (Sanchez & Dunning, 2018). This bubble arises, in part, because beginners develop premature and error-ridden theories about how to approach the task based on very little data (Asch, 1946; Griffin & Tversky, 1992; Jones, Rock, Shaver, Goethals, & Ward, 1968; Kelley, 1950; Sanchez & Dunning, 2018; Tversky & Kahneman, 1971). They reach these conclusions based on small samples of experience because they neglect just how just how noisy and potentially misleading those experiences are.

We examined whether JTC was connected to an enhanced development of this bubble in learning, in part because it exacerbates this latching onto confident theories based on too little data. Consistent with schizophrenia research, we would expect divergent patterns in the beginner's bubble in those displaying high versus low JTC behavior. First, in terms of performance on probabilistic tasks, schizophrenia patients perform worse than do non-clinical participants (Averbeck et al., 2011, Dowd et al., 2016; Weller et al., 2009). However, the rates of learning among schizophrenia patients and control participants are similar or equivalent across most of the existing literature on other types of tasks (Beninger, 2003; Kéri et al., 2000; Weickert et al., 2002; Weiler et al., 2009), ruling out that these effects are solely caused by a lack of attention during these learning tasks. In terms of confidence, schizophrenia patients have also shown to be overconfident in their judgments (Balzan, 2016; Köther, et al., 2012; Moritz & Woodward, 2006a). Therefore, we anticipated that high jumpers would have an elevated confidence curve compared to low jumpers.

Underlying Mechanisms

We also explored cognitive processes that might underlie any relationship of JTC to judgmental errors and biases. In particular, we examined the potential roles played by automatic versus controlled thought, popularly known as system 1 versus system 2 thinking (Kahneman, 2011; Sloman, 1996). For example, there is a way in which JTC can be described as a problem with automatic thought. People showing JTC behavior give greater weight to the stimulus or ideation that happens to be in front of them at the moment they make a judgment. This notion is consistent with the idea that schizophrenia patients have much more difficulty inhibiting sensory information perceived in the immediate environment (Ventura, Wood, & Hellemann, 2013).

Research findings support the perspective that individuals that exhibit JTC behavior imbue current stimuli with "hypersalience." When asked to report confidence of judgments after each draw in the urn task, schizophrenia patients are more confident from the onset (Speechley, Whitman, & Woodward, 2010) and make more drastic changes in their confidence ratings after each additional piece of evidence is seen (Garety, 1991; Huq, Garety, & Hemsley, 1988; Langdon, Ward, Coltheart, 2008) or reward gained (Joyce, Averbeck, Frith, & Shergill, 2013). In short, they overadjust their conclusions based on the last piece of evidence they encounter. This hypersalience account would suggest automatic processing would have a strong impact on misdirected thought in schizophrenia, delusion, and misguided belief. Whatever is present at the moment the individual makes a decision "looms large" in what that decision will be. Automatic thought, thus, "engulfs the field" of an individual's decision process.

In contrast, other research suggests that the issue in JTC may lie in an absence or deficit in controlled processing. More specifically, these accounts assert that JTC follows from a less stringent criterion about how much evidence must be considered and weighed before a decision is reached. JTC is simply a "liberal response bias" or "lower decision threshold," a willingness to reach a conclusion based on little evidence lightly considered. As such, individuals that show high JTC behavior may not have heightened automatic experiences, but instead exert fewer controlled efforts to countermand the impact of any automatic phenomenology. That is, automatic ideation is no more salient. Instead, they fail to engage in cognitive deliberation that would countervail the impact of that automatic ideation.

Research findings give support to this interpretation of lax decision thresholds in cognition among schizophrenia patients. When given many possible hypotheses, individuals showing high JTC behavior rate all possible interpretations of a situation as more plausible than do their low JTC counterparts (Moritz & Woodward, 2004). They can also more often reach decisions when alternative conclusions are still possible (Moritz & Woodward, 2004; Sanford et al., 2014; Woodward, Moritz, Cttler, & Whitman, 2006). Moreover, when schizophrenia patients formulate rules or ideas about how to perform tasks they show high rates of preservative errors, failing to abandon their initial beliefs quickly reached even in the face of disconfirming evidence (Buchy, Woodward, & Liotti, 2007). They are also less likely to change their behavior even when they receive negative feedback (Everett, Lavoie, Gagnon, & Gosselin, 2001), an error not due to lower levels of general cognitive ability (Pantelis, Barber, Barnes, Nelson, Owen, & Robbins, 1999).

We should note that other possible explanations for JTC have been ruled out as plausible explanations according to empirical evidence. JTC is not an impulsivity bias in disguise, nor a deficit in motivation. When the task is harder, those showing both high and low JTC behavior increase the evidence they need before reaching a decision, inconsistent with the idea those who show high JTC behavior are unmotivated to do well or are simply impetuous thinkers (Dudley, John, Young, & Over, 1997; Garety et al., 2005). If the experimenter sets the amount of information to be gathered, those prone to JTC do just as well in reasoning as do other individuals (Dudley et al., 1997) and take equal amounts of time to reach their decisions (Moritz & Woodward, 2005). They also do not act as though additional information is considered more costly, as suggested by motivational theories (Moutoussis, Betall, El-Deredy, & Dayan, 2011).

JTC is also not prompted by need for closure, nor is it due to heightened misunderstandings of Bayesian or probabilistic inference (Dudley et al., 2015). There is emerging evidence suggesting JTC is linked to impairments in working memory, but not to general deficits in cognitive ability (Broome et al., 2007; Garety et al., 2013b). However, that evidence is nascent and inconsistent, in that memory aids do not eliminate these effects (Dudley, John, Young, & Over, 1997; Moritz & Woodward, 2005).

As such, hypersalience and liberal acceptance are the two explanations most supported, with some thought that they are not mutually exclusive and may interact to generate high JTC behavior (Evans, Averback, & Furl, 2015).

Alternative Interpretations

To be sure, we included measures of other individual differences to assess their possible relationship to JTC. We examined two personality differences: need for cognition (Cacioppo & Petty, 1982; Cacioppo, Petty, Kao, & Rodriguez, 1984; Petty, Cacioppo, & Kao, 1984) and need for closure (Kruglanski et al, 2013), that would appear to be possibly related to JTC. We also examined the role played by schizotypy (i.e., subclinical patterns of thought and behavior associated with schizophrenia; Raine & Benishay, 1995), general cognitive ability, and performance in cognitive reflection (Frederick, 2005).

Intervention

Finally, in our last study, we examined whether an intervention designed to prevent JTC behavior among schizophrenia patients would have a similar impact on a nonclinical population. Specifically, we examined whether materials taken from Metacognitive Training (MCT) interventions would reduce JTC within a nonclinical population. MCT is a treatment aimed at correcting cognitive biases associated with hallucinations and delusions in psychosis (Aghotor, Pfueller, Moritz, Weisbrod, & Roesch-Ely, 2010; Balzan, Delfabbro, Galletly, & Woodward, 2014; Favrod et al., 2014; for a review see: Moritz, Andreou, et al., 2014; Moritz, Kerstan, et al., 2011; Moritz, Veckenstedt, et al., 2011; Moritz, Veckenstedt et al., 2014; Moritz et al., 2013; Moritz, Veckenstedt, Randjbar, Vitzthum, & Woodward, 2011). Indeed, MCT interventions have been shown to reduce delusional thinking and its severity, as well as psychotic symptoms, in schizophrenia patients (Garety, Waller et al., 2014; Gaweda et al., 2015; So, et al., 2015).

We specifically asked if MCT would reduce overconfidence as people learned to approach a complex novel probabilistic learning task. Would it prevent the “beginner’s bubble” of overconfidence that people often display as they learn a new task (Sanchez & Dunning, 2018)?

Overview

In sum, the five studies described herein comprised a multi-faceted survey of the relationship between JTC behavior and judgment quality within a nonclinical population. Studies 1a and 1b explored a number of issues in judgment and decision-making. It first explored whether people who show high JTC behavior are more vulnerable to the types of errors inspired by system 1, as well as whether JTC would be associated with agreement with outlandish beliefs, such as conspiracy theories. Further, it looked at “knowledge corruption,” whether people were worse in judging the quality of their decisions. This would be evidenced by overconfidence in judgment, as well as less of a difference between the confidence placed in right answers versus that placed in wrong ones (i.e., confidence discrimination). Finally, it examined the role of automatic and controlled processing in JTC behavior.

Studies 2 through 4 turned to the specific challenge of learning a complex task—predicting uncertain events—and more directly to the issue of the development of overconfidence among beginners. The specific task was diagnosing illnesses from medical symptoms. As such, we could examine the role played by JTC in learning a skill versus assessing confidence in that skill, that is, how one’s subjective learning curve differs from one’s objective learning curve. Study 2 examined whether JTC predicted differences in confidence and performance in the task, and the development of a “beginner’s bubble” of overconfidence (Sanchez & Dunning, 2018). Study 3 examined how differences found could be traced to overly exuberant but faulty strategies that people adopt quite early in the learning process. Study 4 examined whether training people not to jump to conclusions—an intervention modeled after those used in schizophrenia treatment—would reduce any overconfidence we observed.[12]

Studies 1a and 1b: Error, Belief, and Knowledge Corruption

Studies 1a and 1b were designed as a general survey of the relation between JTC and judgmental quality in a nonclinical population. This study was conducted to determine if findings from schizophrenia research would echo within a nonclinical population. Previous research has touched on questions like this. Higher levels of JTC have been related to increased paranoid thoughts and the occurrence of perceptional anomalies within the general population (Freeman, Pugh, & Garety, 2008). High JTC behavior has also been related to a greater belief in conspiracy theories (Moulding et al., 2016). However, we wanted to see if JTC was connected more generally to problems in judgment and decision-making in a nonclinical population. Specifically, we measured those who tended to jump to conclusions in a task, and then gathered their responses across several different judgment tasks, expecting that high jumpers would make more errors and exhibit more bias.

More specifically, we tested first how vulnerable people showing high JTC behavior were to system 1 biases. We were interested in examining the relationship between JTC and system errors because the proposed mechanisms, liberal acceptance and hypersalience of evidence, from schizophrenia research were reminiscent of them and appear to represent errors in logical reasoning. In regards to system 1 errors, we tested whether high jumpers showed flawed logical reasoning and analytical skills using three methods, the Cognitive Reflection Test (CRT),

[12] Studies 1a and 1b were preceded by three exploratory pilot studies testing the relation of JTC to these same biases. All these exploratory studies produced effects that equaled or exceeded the strength of associations we report herein. As such, Studies 1a and 1b constituted confirmatory studies following-up on these pilot results. Studies 2-4 constitute all studies we have conducted on JTC and probabilistic learning. They were all confirmatory in nature.

a belief bias test, and denominator neglect. The CRT measures whether a common quick but wrong answer can be resisted in order to arrive at a correct answer (Frederick, 2005). The belief bias test measures logical reasoning, in which analytical reasoning is pitted against world knowledge. Belief bias occurs when people are led to a logical error by endorsing the conclusion consistent with their world knowledge. That is, real world beliefs have been shown to be a system 1 obstacle to correct logical analysis (Evans & Curtis-Holmes, 2005). In our final test of logical reasoning, we examined denominator neglect, a common error made in risky choice. Specifically, in gambling decisions, people often err by paying attention to the numerator of a probability but neglecting its denominator. Previous work suggests that such errors arise out of intuitive, quick thought (Denes-Raj & Epstein, & Cole, 1995).

We also conducted other analyses aimed at specifying the ingredient within JTC behavior that disrupts the quality of judgments. We tested whether it was enhanced reliance on automatic ideation or instead a lack of controlled processing that related to lower quality judgments in those exhibiting higher JTC behavior. Using belief bias and denominator neglect, we assessed how much participants gave weight to automatic versus controlled processes in their decision making.

Using Jacoby's (1991) process dissociation scheme, we took questions from the belief bias tests and the denominator neglect task and parsed out whether they associated with automatic (i.e., real world knowledge) or controlled processes (i.e., logical reasoning). For both tasks, we included items in which automatic and controlled processing supported the same correct answer (consistent items) versus those in which controlled thinking supported the right answer and automatic processes the opposing wrong one (conflict items). We then used these two categories of questions to calculate how much weight participants gave to automatic versus controlled thinking. The controlled component, or C, of a participant's judgment is given by the equation:

$$C = P(\text{correct}|\text{consistent items}) – P(\text{incorrect}|\text{conflict items})$$

The automatic component, or A, is given by:

$$A = P(\text{incorrect}|\text{conflict items})/(1\text{-}C)$$

Second, we tested whether those exhibiting high JTC behavior are more likely to endorse oddball beliefs (medical myths, conspiracy theories, and paranormal beliefs) and demonstrate overconfidence. In schizophrenia research, high jumpers have been found to hold more oddball beliefs and are more likely to make more overconfident errors. To measure overconfidence, participants completed a 20-item quiz about American history and civics. For each answer they gave, they estimated the chance that it was correct. From those estimates, we could construct two indices of knowledge corruption. First, how much would participants overestimate the chances that their answers were correct (i.e., overconfidence)? Second, we could gauge how informative their confidence estimates were—that is, to what degree did their confidence estimates differ between answers they got right versus wrong (i.e., discrimination)? We predicted that those that exhibited high JTC behavior would display more of the former (overconfidence), but less of the latter (discrimination).

Throughout our analyses we also examined potential confounding individual differences (such as the need for cognition or schizotypy) to see if they played any role in our judgment outcomes.

Method

We describe the methods and results of Studies 1a and 1b together, given their similarity. Study 1b was a preregistered replication. Details can be found on the Open Science Framework, https://osf.io/qmd8w/.

Ethics Statement. Studies 1a, 1b, 2, and 3 were judged exempt from human subjects approval from the Institutional Review Board of Cornell University. The last study was approved by the IRB under the title Metacognitive Training (MCT) to Reduce Judgment Error (Protocol #1503005461).

Participants. In Study 1a, participants were 289 individuals recruited from Amazon's Mechanical Turk crowd sourcing facility (54% male; 46% female). Their participation spanned two sessions. They were paid $5 for participation for the first. Approximately a week after completing it, participants were invited to complete the second session for $3. A total of 257 respondents did. Data from an additional 18 participants were omitted because they had participated in pilot studies leading to this data collection. We had originally aimed at sample size of at least 260 to provide a 90% of capturing a correlation of size .20, but allowed the sample to grow larger until the end of the day's business hours when actually running.

In Study 1b, we aimed to recruit 350 individuals from Amazon's Mechanical Turk crowd sourcing facility. Due to the vagaries of Amazon's Mechanical Turk, 346 participants completed the hit. Our sample size was larger than Study 1a because we planned to exclude participants that failed at least one attention check. We again anticipated that a sample size of at least 260 would provide a 90% of capturing a correlation of size .20. After omitting 48 participants for failing attention checks, we were left with a final sample of 298 (59% male; 41% female). Participants were paid $5 for participation for completing the study.

Procedure and Measures. In Study 1a, participants addressed a variety of surveys and tasks over two sessions. During the first session, participants completed the following tasks in the order described here. In Study 1b, participants completed the same surveys and tasks, in the same order, albeit in one session.

***Jumping to Conclusions*.** First, participants completed a task that assessed jumping to conclusions (JTC). Participants confronted a scenario in which they observed a character fishing from one of two lakes. One lake contained an 80% majority of red fish, with the rest being grey. The other contained an 80% majority of grey fish, with the remaining minority being red. The participant observed the color of each individual fish the character caught, one at a time, until that participant was ready to decide whether the character was fishing out of the majority-red or majority-grey lake. The participant could reach their decision at any time, including after just seeing the first fish the character caught, or could ask to see the next fish. Once the participant reached a decision, the task was over. Otherwise, the task terminated after the eighth fish and participants were asked to reach a decision. After completing this task, participants repeated it with a new character and pair of lakes, in which the two lakes contained either a 65% majority of red or grey fish. The same rules applied as above.

On average, participants asked to observe 3.2 fish per lake (SD = 1.94; range 1 to 8 fish) in Study 1a and 3.55 per lake (SD = 2.11, range 1 to 8 fish) in Study 1b. To assess JTC, we summed the number of fish they asked to see across the two lakes, with the number they requested quite highly correlated across the two lakes, $r(287) = .81$, $p < .001$, in Study 1a and $r(296) = .79$, $p < .001$, in Study 1b. To reduce potential skew and kurtosis, we took the natural log of this sum. Then we reverse coded this number by subtracting the log of the maximum amount of fish that could be viewed by the natural log of the number of fish viewed (2.772589 – the natural log of the sum of the fish). We should note that 34% and 26% of respondents in Studies 1a and 1b, respectively, (ns = 98 and 78) asked for a maximum of two fish across both lakes, often considered the specific behavioral signature of JTC behavior (Dudley et al., 2015), indicating its presence in a nonclinical population.

Oddball beliefs. Participants then indicated their agreement with a series of oddball beliefs. First, they answered six questions on medical myths taken from Oliver & Wood (2014). Example items include *"Health officials know that cell phones cause cancer but are doing nothing to stop it because large corporations won't let them"* and "*Public water fluoridation is really just a secret way for chemical companies to dump the dangerous byproducts of phosphate mines into the environment*." Second, participants reported their agreement with eight traditional conspiracy theories (e.g., *"The Apollo moon landings never happened and were staged in a Hollywood film studio,"* and "*Princess Diana's death was not an accident but rather an organized assassination by members of the British royal family who disliked her*") taken from Lewandowsky, Gignac, & Oberauer (2013). Finally, participants rated their agreement with eleven statements about paranormal phenomena, such as "*During altered states, such as sleep or trances, the spirit can leave the body*," or "*A person's thoughts can influence the movement of a physical object,*" taken from the Revised Paranormal Belief Scale (Tobacyk, 2004). We included those statements that seemed the most outlandish as well as avoided any reference to religion.

Participants rated their agreement along 5-point scales that were anchored "strongly disagree" to "strongly agree" at either end. Mean agreement in Study 1a was 2.02 (SD = 0.96) for medical myths, 1.95 (SD = 0.85) for conspiracy theories, and 2.10 (SD = 0.77) for paranormal beliefs. In Study 1b, the comparable means were 2.07 (SD = 0.99), 2.06 (SD = 0.91), and 2.13 (SD = .87), respectively. The proportion of participants who scored above the scale midpoint (and thus tipped generally toward belief rather than disbelief) were 13.5%, 19%, and 13%, respectively, in Study 1a; 18%, 16%, and 17% in Study 1b.

Overconfidence. Participants then completed the overconfidence task. They were asked

20 different questions about American history and civics, such as *how many senators are there* and "*the Bill of Rights is.*" Participants were given 4 response options from which to choose for each question. They were also asked, for each item, the chance that their answer was right from 25% (I am only guessing) to 100% (I am certain that I am right). For each answer they gave, they estimated the chance that it was correct. From those estimates, we could construct two indices of knowledge corruption. First, how much would participants overestimate the chances that their answers were correct. Second, we could gauge how informative their confidence estimates were—that is, to what degree did their confidence estimates differ between answers they got right versus wrong? How well would their confidence estimates, thus, discriminate between right or wrong answers? We predicted that those exhibiting high JTC behavior would display more overconfidence, but their estimates would show less discrimination.

Across the quiz, participants on average achieved a score of 73.4% (SD = 17.5) correct in Study 1a and 77.6% (SD = 15.2) in Study 1b, but expressed 83.2% (SD = 12.9) confidence in their answers in Study 1a and 85.8% (SD = 11.1) in Study 1b. As such, as a group, they were overconfident in both studies: Study 1a, $t(288) = 11.36$, $p < .001$, $d = 1.30$; Study 1b, $t(297) = 9.42$, $p < .001$, $d = 1.09$.

Supplemental Scales. Participants then completed questionnaires relevant to other research streams, but toward the end of the session completed four scales pertinent to the research reported here. First, to assess general cognitive ability, participants completed a 50-item excerpt of the Ammons Quick Test of intelligence (Ammons & Ammons, 1962). In the excerpt, participants were shown ten words and, for each, asked which of four different drawings was most relevant to it. In all, participants judged 10 words associated with five different panels of drawings ($n = 279$, $M = 37.5$, SD = 10.0 in Study 1a, $n = 298$, $M = 39.96$, SD = 7.65 in Study

1b). The Ammons test correlates positively with the WAIS verbal scale (Cull & Colvin, 1970).

Second, participants completed the SPQ-B, a brief 22-item version of the instrument that screens for schizotypal disorder (*n* = 278, *M* = 8.0, SD = 5.5, for Study 1a, *n* = 298, *M* = 7.9, SD = 5.3 in Study 1b; Raine & Benishay, 1995), the need for cognition scale (*n* = 278, *M* = 80.7, SD = 22.5, in Study 1a; *n* = 298, *M* = 87.0, SD = 32.2 in Study 1b; Cacioppo, Petty, Kao, & Rodriguez, 1986), and the need for closure scale (*n* = 278, *M* = 156.4, SD = 18.2 for Study 1a; *n* = 298, *M* = 175.4, SD = 25.5 in Study 1b; Webster & Kruglanski, 1994).

System 1 Biases. Participants completed three tasks that measured the ability to resist errors inspired by system 1 processes.

Belief Bias. Participants were asked to evaluate the logical validity of 24 different syllogisms and told to ignore real world associations. They specifically stated whether each syllogism was valid (i.e., the conclusion followed from the preceding premises) or was invalid (i.e., the syllogism was untrue or one could not tell). Of key interest were 8 items that tested for belief bias, where logical validity and real-world associations were put in conflict (e.g., *Flowers need water. Roses need water. Therefore, roses are flowers.*) (Evans & Curtis-Holmes, 2005). On these items, participants made an average of 3.5 errors, *SD* = 2.49, *n* = 257, in Study 1a and 3.4 errors in Study1b, *SD* = 2.72, *n* = 298. Of the 16 other syllogisms, 8 contained conclusions supported both by real world associations and logical validity, e.g., *"All fish swim. Tuna are fish. Therefore tuna can swim."* The final 8 items were designed to avoid real world associations all together, e.g., *"All Posed are red. The item in my hand is red. Therefore, the item in my hand is a Posed."*

Cognitive Reflection Test. Second, participants completed the 7-item version of the Cognitive Reflection Test (CRT; Toplak, West, & Stanovich, 2014). The CRT (Frederick, 2005) measures whether people can resist a common quick but wrong answer in order to arrive at a correct answer. For example, *"If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?"* The quick but wrong intuitive answer that typically comes to mind is 9 days, but the true answer is 4. On the CRT, participants made an average of 3.9 errors, *SD* = 2.17, *n* = 242 in Study 1; 3.8 errors, SD = 1.70, in Study 1b.

Denominator Neglect. We presented participants with several gambles involving a choice between gambling on an Urn A or B. We tested errors made by paying attention to the numerator of a probability but neglecting its denominator. For example, people may prefer a gamble they have an 8 out of 100 chance of winning over one where they have a 1 in 10 chance of doing so. That is, they pay attention to the numerators of both gambles (8 versus 1) but neglect the denominators (100 versus 10). Thus, they fail to recognize that they have a better chance of winning on the second gamble (10%) than the first (8%) (Denes-Raj & Epstein, & Cole, 1995).

In the task we used, with Urn A, they always had a 1 in 10 chance of winning (i.e., there was one winning ball in the urn and 9 losing balls). Urn B had 100 total balls, but the number of winning balls varied. Participants were asked how much they preferred Urn A or B when Urn B had 7, 8, 9, 10, 11, 12 or 13 winning balls. Participants responded on a 7-point scale that ranged from 1 (definitely Urn A) to 7 (Definitely Urn B).

To assess denominator neglect, we averaged participants' preferences for those three judgments where the Urn A provided a better chance of winning but Urn B presented a larger numerator (i.e., number of balls that could win were 7, 8, or 9). The average preference rating was 2.38, *SD* = 1.48 in Study 1a and 2.67, *SD* = 1.72, in Study 1b, showing an overall preference for the correct urn. However, 19% of participants on average were indifferent or rated the other urn as the one they preferred in Study 1a, as did 26% in Study 1b.

Revisions in Study 1b. Study 1a and 1b were identical with two exceptions.

Confidence in Jumping-to-Conclusions Task. One additional set of measures was included in the research plan: In the fish-lake task, after the color of each fish was reported to participants, they answered the following question: *What do you think is the chance that the fisherman is at Lake Gamon or Lake Rekon? (Your percentages should add to 100%).* They then reported the chance the fisherman was fishing from both Lake Gamon and from Lake Rekon. They were then asked which lake the fisherman was fishing from or whether they would like to see the next fish, as in Study 1a. The same procedure was done for the second lake. From this measure, we can take participants initial confidence upon seeing the first piece of evidence in the fish-lake task, as well as their final confidence upon making a decision about which lake the fish were taken from.

Attention Checks. Participants also completed two attention checks. The attention check was conducted after the civics questions. The attention check instructed participants to provide a best guess for the years containing a series of historical event, then a lower-bound year and finally an upper-bound year for their estimates, such that the chance that their bounds contained the event's true year was 80%. They then progressed to the next screen, which simply displayed *Thomas Edison invents the light bulb,* followed by three text boxes. Those leaving the boxes empty or filling them with irrelevant text were omitted. Participants were excluded who gave a lower bound estimate greater than or equal to the best guess or a best guess equal to or greater to the higher bound estimate. The same procedure was followed for the next screen, which merely displayed the event of the stock market at Wall Street crashed, starting the Great Depression.

Results

Recall that, for clarity of presentation, scores on the JTC scale in all studies were reversed such that high scores indicated a greater degree of jumping to conclusions. Scores on CRT, belief bias, and denominator neglect were similarly scored in each study such that higher numbers represented more error. Table 1 depicts the zero-order correlations among all measures taken in Study 1a and 1b.

Error, Belief, and Overconfidence

As Table 2 shows, JTC behavior was significantly correlated with all judgmental errors tested. For example, after standardizing scores on system 1 bias measures (i.e., belief bias, CRT performance, and denominator neglect) so that they shared a common scale, a MANOVA showed that JTC was significantly and positively correlated with their expression, $F(1, 287) = 34.68$, $p < .001$, $\eta_p^2 = .36$ in Study 1a and $F(1, 296) = 69.84$, $p < .001$, $\eta_p^2 = .44$, in Study 1b. A similar MANOVA showed that JTC was positively associated with endorsements of oddball beliefs in both studies: Study 1, $F(1, 287) = 14.58$, $p < .001$, $\eta_p^2 = .22$; Study 1b, $F(1, 296) = 32.76$, $p < .001$, $\eta_p^2 = .22$; Study 1b, $F(1, 296) = 32.76$, $p < .001$, $\eta_p^2 = .33$. As Table 2 shows, JTC was also significantly related to each system 1 bias individually, as well as each separate category of oddball belief.

In terms of a direct test of overconfidence, although JTC was related to marginally less confidence in answers to the civics test in Study 1a but not Study 1b, *p*s = .064 and .354, respectively, it also led much less accuracy in both studies, both *p*s < .001. The net effect was those that were more likely to jump to conclusions in the JTC task were more overconfident (i.e., confidence minus accuracy) in their answers than those that were less likely to jump to conclusions in the task, both *p*s < .001 (see Table 2). Figure 1 (left panel) depicts the level of confidence, accuracy, and overconfidence displayed by participants high in JTC (+1 SD) versus low (-1SD) in each study. As seen in the figure, high JTC participants show a greater gap between confidence and accuracy, particularly because they achieve lower accuracy in their answers.

Lastly, we examined how well participants' confidence judgments discriminated between times they were right versus when they were wrong. As such, for each participant, we took the average of confidence estimates for accurate answers on the civics quiz, and separately did the same for their inaccurate answers. The question is how much confidence for accurate answers exceeded confidence for inaccurate ones. However, we did not simply take the difference in confidence between the two circumstances because 13 participants in Study 1a and 10 in Study 1b achieved perfect scores and thus could provide no difference score. Thus, as an alternative, we conducted a mixed model analysis in which we examined participants' accurate and inaccurate averages, their JTC level, and the interaction between JTC and accuracy with participant as a random variable (random intercept and slope). Such an analysis can accommodate missing data, and under such we would predict an JTC X accuracy interaction, which we observed, $F(1, 274.0) = 6.47$, $p = .012$, $\eta_p^2 = .15$, in Study 1a and $F(1, 286.3) = 7.60$, $p = .006$, $\eta_p^2 = .16$, in Study 1b. As Figure 1 (right panel) shows, the gap in confidence between right and wrong answers was smaller for high JTC participants than for their low JTC peers.

In sum, all findings were consistent with predictions and with findings from the schizophrenia literature. People exhibiting high JTC behavior in a nonclinical population were more likely to commit system 1 errors, endorse oddball beliefs, and demonstrate more knowledge corruption, in that the confidence they imbued in their answers was less indicative on two measures (overconfidence, discrimination) of the actual accuracy of those answers.

Alternative Measures

In our second set of analyses, we examined whether other individual difference measures could serve as potential confounds to JTC behavior. Table 2 summarizes the relationship between JTC and our dependent measures once various other individual difference measures are controlled for. In sum, controlling for schizotypy, need for cognition, and need for closure did nothing to diminish the relationships between JTC and our central dependent measures.

Controlling for CRT and general cognitive ability did diminish the relationship between JTC and our dependent measures, but JTC still predicted significant unique variance in our outcome measures. Specifically, it is not a surprise that controlling for CRT would diminish the association between JTC and our outcome measures, because we predicted that JTC and CRT would be related. That said, JTC was still related to system 1 errors (i.e., belief bias and denominator neglect), oddball beliefs, and overconfidence, *p*s = .003, .043, and < .001, respectively, in Study 1a, and *p*s = .002, < .001, and < .001, respectively, in Study 1b, even after controlling for CRT.

The same results arose after controlling for general cognitive ability: The relationship between JTC and outcome measures was reduced but still revealed a unique significant relationship to system 1 errors, *p*s < .001 for both Studies 1a and 1b, oddball beliefs, $p < .001$ and = .007, for Study 1a and 1b, respectively, and overconfidence, $p = .012$ and .001, for Study 1a and 1b, respectively.

In terms of individual outcome variables, the relationship between JTC and paranormal beliefs, as well as denominator neglect, became nonsignificant after controlling for CRT and general cognitive ability in Study 1a, but remained significant in Study 1b. The relationship between JTC and discrimination became marginally significant after controlling for CRT in

Study 1a ($p = .088$) but remained significant in Study 1b ($p < .001$). However, in the main, the relationships between JTC and our outcome variables generally remained significant—significantly so when we meta-analytically combined data across the two studies, all zs > 3.13, ps < .001.

Mechanism: Automatic v. Controlled Processes

We next took participants' performance on belief bias and denominator neglect tasks and broke them down into estimates of their automatic and controlled components. The estimates revealed that the automatic component to the belief bias task was quite strong relative to the controlled component, Ms = .63 and .38 for A and C, respectively, in Study 1a and .52 and .35, respectively, in Study 1b. The denominator neglect task showed the exact opposite, Ms = .25 and .63 for A and C, respectively, in Study 1a, and .37 and .55, respectively, in Study 1b.

The estimates of controlled processing derived from the two tasks correlated highly with each other, $r(247) = .43$, $p < .001$, in Study 1a and $r(295) = .54$, $p < .001$ in Study 1b; the two automatic components, however, were not as strongly correlated, $r(247) = .09$, $p = .148$ in Study 1a and $r(295) = .23$, $p < .001$, in Study 1b—reminiscent of other studies that have found little correlation between automatic measures (Bosson et al., 2000; Weinberger & McClelland, 1990; Winter, 1999).

How much was JTC behavior related to automatic and controlled processes (A and C) in judgment? To assess the relationship, we constructed a regression analysis predicting JTC from automatic and controlled components for belief bias and denominator neglect separately. For weights derived from the belief bias task, the controlled component correlated strongly with JTC behavior, βs = -.38 and -.37, η_p^2s = .11 and .08, ps < .001, for Study 1a and 1b, respectively, but the automatic component did not, $\beta = -.12$, $p = .064$, η_p^2s = .01, and $\beta = -.03$, $p = .638$, η_p^2s = .00, for Study 1a and 1b, respectively. For estimates derived from the denominator neglect task, JTC again was predicted by the controlled component of judgment, βs = -.25 and -.31, η_p^2s = .05 and .08, *p*s < .001, for Study 1a and 1b, respectively, but not the automatic component, β = .02 and .10, η_p^2s = .00 and .01, both *ns*, for Study 1a and 1b, respectively. Figure 2 depicts the relationship between C and A, derived by both methods, to JTC.

This association of JTC to reduced controlled processing explained its relationship to our outcome measures of error, false belief, and overconfidence. Table 3 depicted how strongly *C* and *A*, derived from either task, predicted judgment quality and error. As seen in the table, with only one exception, *C* predicted decision performance on all outcome measures, regardless of which task it was derived from. *A* failed to predict performance on any outcome measure.

In short, our data strongly suggest an account in which JTC predicts judgment performance because it marks a reduction in controlled processing. Indeed, if we conduct the final two tests of a strict mediational analysis, we largely find a consistent story. Specifically, first, as Table 4 shows, after controlling for JTC, *C* continues to predict judgment performance on all outcome variables, regardless of which task it is derived from, with two exceptions (C from the denominator neglect task predicting confidence and discrimination). This provides evidence of mediation. Second, with only four exceptions, the relationship we find evidence when we test for them of indirect effects that pass through *C,* using the *lavaan package* from *R*, estimating 95% confidence intervals by bootstrapping samples with 1000 iterations.

In addition, and consistent with this link between JTC and reduced analytic thought is our analysis of confidence as participants completed the fishing task. Did high JTC participants reach their decisions on the task so quickly because they were more persuaded by initial pieces of evidence, consistent with a hypersalience hypothesis, or were they instead willing to reach a decision at lower levels of confidence, as suggested by the low threshold hypothesis? To test both hypotheses, we examined the average initial confidence participants expressed upon seeing the first fish in the task for each lake. We also examined the final confidence participants expressed at the time they reached a decision for each lake. This analysis provided evidence for both hypotheses. That is, initial confidence upon seeing the first fish was positively correlated with JTC, $r(295) = .20$, $p < .001$. However, and to a stronger degree, JTC was associated with declaring a final decision at lower levels of confidence, $r(25) = -.40$, $p < .001$ (see Figure 3).

Thus, those who exhibit high JTC behavior were more impressed with initial pieces of evidence and also were more willing to commit to a decision at lower thresholds of confidence. However, further analyses suggested this was an independent outcome of JTC, different from the lack of analytic thought demonstrated earlier. That is, mediational analyses exploring confidence during the JTC task failed to mediate the absence of *C* in JTC thought, and the absence of *C* failed to mediate the link between JTC and confidence while completing the fish-lake task.

Finally, data tended to contradict a more mundane explanation for JTC behavior suggesting that high JTC behavior reflects being impulsive or unmotivated to do well. Specifically, both high and low jumpers were influenced similarly in both studies by the ambiguity in the fishing task, replicating past results (Dudley et al., 1997; Garety et al., 2005). Recall that in both studies that the second lake we presented participants represented a more ambiguous task than the first because the majority of fish in the lakes was smaller (65%) than the first (80%). Rationally, participants should select more fish from the second lake before making a call than from the first lake, and this is what we observed. On average, participants asked to see 0.60 more fish, $SD = 1.32$, $t(288) = 7.76$, $p < .001$, $d = 91$, in Study 1a and 0.61 more fish, $t(297) = 7.70$, $SD = 1.37$, $p < .001$, $d = .89$, in Study 1b. This tendency was only minimally correlated with overall number of fish each participant drew, *r*s = .07, *ns,* and .16, $p = .005$, for Studies 1a and 1b, respectively. More important, a regression analysis suggests that this difference in effort remains significant if we focus on the minimum number (2) of fish that participants could draw, 0.48, $t(287) = 4.09$, $p < .001$, $d = .48$, and 0.33, $t(296) = 2.54$, $p = .012$, $d = .30$, for Studies 1a and 1b, respectively.

Study 2: Confidence Versus Accuracy in Learning

In the previous studies we found that the jumping to conclusions is associated with a number of flaws and biases in everyday judgment, including overconfidence. We decided to extend our investigation by looking at overconfidence in one particular circumstance: When people learn a new complex skill.

In Study 2, participants completed a probabilistic learning task used in previous research (Sanchez & Dunning, 2018). In this task, participants are asked to imagine they are medical residents in a post-apocalyptic world that has been overrun by zombies. They must diagnose 60 different patients who were at risk for zombie infection. There were two different zombie diseases to identify and, of course, patients who could be healthy. Participants were given information on eight different symptoms in each case (e.g., brain inflammation, fever, glossy eyes). After each patient, they received feedback about their accuracy.

Consistent with prior research, we predicted that participants overall would achieve incremental increasing success in learning how to diagnose patients. Their confidence, on the other hand, would follow a very different path, blowing up into a “beginner’s bubble” (Sanchez & Dunning, 2018). That is, participants would start largely and appropriately calibrated in their first few judgments, showing confidence that matched their levels of accuracy, but with confidence racing ahead of accuracy after only a little experience at the task. This bubble of overconfidence would then deflate somewhat as participants realized that their initial burst of confidence was somewhat inappropriate. Finally, near the end of the 60 learning trials, confidence would resume its upward climb.

However, we anticipated that the propensity to form this beginner's bubble of overconfidence would be correlated with JTC behavior. We anticipated that high jumpers would perform worse than low jumpers because they would rely on less evidence to reach conclusions about how to approach the task, and these premature conclusions would be detrimental to their overall performance. Their confidence, however, would follow a different pattern. We predicted the same curvilinear trajectory for high and low jumpers but hypothesized that the overconfidence bubble for high jumpers would be relatively exaggerated by their premature settling on an approach to the task. Taken together, we expected high jumpers would perform worse, but exhibit greater confidence across all their decisions, thus ending up more overconfident than low jumpers.

Method

Participants. One hundred participants were recruited from Amazon's Mechanical Turk crowdsourcing facility. This study was conducted in two parts over two days. Participants received \$5 for their participation of part one. For completion of part two, participants received \$3 and had the chance to win an additional \$3 if they achieved an overall accuracy level of 80% in the medical diagnosis task described below. Three participants were excluded from the analysis because they failed to complete part 2. The sample consisted of 46% men and 54% women.

To enhance statistical power, we conducted a within-subject design to control for error due to participant. Given these techniques, we used a rather crude estimation procedure to compute our sample size due to uncertainties we faced in the sizes of our predicted effects and complexities of calculating power in the specific data analysis strategy we adopted (Hayes, 2006). We anticipated that our effects, all within-subject, would be moderate in size (d = .25), given pilot data, and so calculated the sample size needed to capture such an effect in a within-subject comparison. A sample size of 86 is required for a statistical power of 80% to detect an effect of d=.25 with $\alpha = .05$, but rounded up our initial sample size to 100 participants to be conservative.

Procedure. Participants were instructed that they would be taking part in a two-day study. On the first day they would be shown several logic tests. They also completed the same fishing task we used in Study 1 as a measure of their JTC behavior, as well as a portion of the Ammons Quick Intelligence Test (Ammons & Ammons, 1962), need for closure scale (Webster & Kruglanski, 1994), and need for cognition scale (Cacioppo, Petty, Kao, & Rodriguez, 1986). JTC was measured using the identical procedure as in Studies 1a and 1b.[13]

Probabilistic Learning Task. The next day, participants completed the second part of the study. In this task participants took part in a probabilistic learning task that has been used in previous research (Sanchez & Dunning, 2018). They were told they would be taking part in a hypothetical medical diagnosis scenario. Two strains of zombie disease had broken out across the world, TS-19 and Mad Zombie Disorder (MZD). Participants were instructed they were being trained in zombie disease detection and treatment. As part of their training they were about

[13] Due to the smaller sample in Studies 2 through 4, we applied outlier screens, using the nonparametric procedure described by Leys, Ley, Klein, Bernard, & Licata (2013), that identifies outliers using the median and median absolute deviation. This procedure in small samples is much more robust against any influence of outliers on the standard deviation of the sample. We adopted the conservative criterion of $k = 3$. Using this procedure identified 3 outliers on the Ammons and the need for closure scales, and 1 on the need for cognition measure, whose data were omitted on the relevant measure. All analyses yielded identical results even if these outliers are included in analyses.

Throughout all of our studies, we also excluded participants who never varied in their confidence judgments on the probabilisitic learning task. For this study, all participants varied in confidence judgments.

to see and diagnose patients. Participants were then presented sixty patient profiles, one at a time. Each profile listed eight symptoms and stated whether each symptom was present or absent in the current patient. Participants diagnosed each patient as having TS-19, MZD, or neither. Two of the symptoms were correlated probabilistically with TS-19, two with MZD, one with both, and three with neither. Participants were warned that symptoms that were connected to disease or health only to a probabilistic degree.

They also reported how confident they were of their decision would prove accurate. Specifically, they were instructed:

> Please report how confident you are in this decision. What's the chance that you are right, from 33% to 100%? Mark 33% if you think it's just as likely that you are wrong as you are right (i.e., it's 33-33-33 that I'm right). Mark 100% if you are absolutely sure that you are right; there's no chance that you are wrong. Mark 66% if you think the chance that you are right is 2 out of 3. Mark whichever probability best indicates the specific chance that you are right.

After participants reported their confidence for each case, they were given immediate feedback on their performance. Feedback included the right diagnosis, and repeated the symptom profile presented for that patient. At the end of the task participants rated how difficult they perceived the probabilistic task to be along a 9-point scale, from 1 (not at all) to 9 (extremely).

Results and Discussion

See Table 5 for all zero-order correlations between measures in Study 2.

Learning and Performance. We first examined whether participants in fact learned as they gained more experience in the medical diagnosis task. We conducted a logistic mixed model analysis (random-intercept, random-slope) assigning experience (i.e., trial number) and JTC score as fixed variables while including participant as a random variable. Accuracy increased on average across the 60 diagnoses participants made, $b = .005$, $se_b = .0016$, $p = .001$, $OR = 1.005$, but high jumpers displayed worse overall accuracy in the task, $b = -.16$, $se_b = .06$, $p = .01$, $OR = .85$. We explored and found no interaction between JTC and experience, $b = -.0002$, $se_b = .002$, $p = ns$. In sum, participants showing high JTC behavior achieved lower levels of performance, but the rate at which they learned was not significantly different from those showing low JTC behavior.[14]

Confidence and Overconfidence. The analysis above suggests that accurate performance rises slowly and incrementally with experience. We expected confidence to show a different and more complex relationship with experience. As an example, we found overall that participants proved to be more confident in their judgments, believing they would be right 63.2% ($SD = 16.1$) of the time when their accuracy rate was only 57.7% ($SD = 9.8$), $t(96) = 2.96$, $p =.004$, $d = .60$. In short, they were overconfident. As seen in Table 5, JTC behavior positively correlated with overconfidence, $r(95) = .34$, $p < .001$, as it did positively with confidence, $r(96) = .23$, $p = .024$, and negatively with accuracy, $r(95) = -.25$, $p = .01$.

More importantly, we predicted that confidence would follow a cubic trend over time on the task, as seen in Sanchez and Dunning (2018). Thus, we first conducted a mixed model analysis on confidence (random intercept and slope) testing whether confidence followed linear, quadratic, and cubic trends across experience, including JTC and its interactions with those trends as independent variables. As expected (see Table 6), the cubic trend in confidence was

[13] We then tested for a quadratic trend in learning by adding a quadratic term for experience. The quadratic term was significant but did not yield a better fitting model, ΔBIC = +14. This was consistent with prior research showing no quadratic trend in learning in this task (Sanchez & Dunning, 2018). People learn to predict outcomes that are uncertain gradually.

significant, b = .0004, se_b = .00007, p < .001, as was the quadratic trend in confidence, b = -.01, se_b = .001, p < .001. JTC mattered as well, as seen in Figure 4, which depicts actual versus perceived performance over the course of the diagnosis task for high (+1 SD above the average) and low JTC participants (-1 SD). Higher JTC was related to greater confidence, b = 5.84, se_b = 2.54, p = .024, and JTC interacted with the quadratic trend line of experience, b = .005, se_b = .002, p = .038. As seen in the right panel of Figure 4, high JTC participants started overconfident in their diagnoses and experienced a further burst of overconfidence in just a few trials. Their confidence then stagnated and fell somewhat, but began climbing toward the end of the diagnosis task. Low JTC participants started starkly underconfident, but their confidence quickly surged—but only to a level that roughly matched their accuracy rate and stayed there, although falling a little. In sum, they ended the task roughly well-calibrated in their confidence, in distinction to their high JTC counterparts.

Finally, we created a measure of overconfidence for each response that participants gave. Accuracy was coded as 100 when the participant was right and 0 when he or she was wrong. We then subtracted that figure from their level of confidence. We then subjected these overconfidence scores to a fuller mixed model analysis (random intercept and slope) including JTC as well as the linear, quadratic, cubic components of experience, and interaction effects between JTC and all experience components as fixed variables and participant as a random variable. There was a significant effect of JTC behavior, b = 9.79, se_b = 2.78, p <.001, η_p^2 = .11, a JTC X linear component interaction with experience (see the left panel of Figure 4), b = -.43, se_b = .15, p = .005, η_p^2 = .08, and a JTC X cubic interaction, b = .0006, se_b = .0003, p = .022, η_p^2 = .04. In sum, and apparent in the right panel of Figure 4, which takes the confidence trends for high and low JTC participants and subtracts the trend shown in accuracy, high jumpers were more overconfident than low jumpers, but that difference diminished somewhat over experience, from about a 30% difference to roughly a 20% one.

Supplemental Measures. Next, we examined whether general cognitive ability, need for cognition, and need for closure qualified the relationship between overconfidence and JTC behavior. We did so by correlating each variable with confidence, accuracy, and overconfidence across participants. Only cognitive ability mattered, in that greater ability was associated with less confidence, $r(92) = -.24$, $p = .004$, and lower overconfidence, $r(92) = -.29$, $p = .004$. Thus, we added it to a multi-level analysis with JTC. With cognitive ability added, the relationship between JTC and confidence became nonsignificant, $b = 3.77$, $se_b = 2.69$, $p = .16$, but the relationship between cognitive ability and confidence also failed to reach conventional levels of significance, $b = -.57$, $se_b = .32$, $p = .08$. For overconfidence, both JTC, $b = 5.66$, $se_b = 2.84$, $p = .049$, $\eta_p^2 = .04$, and cognitive ability, $b = -.94$, $se_b = .40$, $p = .02$, $\eta_p^2 = .06$, were significant predictors. JTC also interacted with the linear trend in experience, $b = -.39$, $se_b = .15$, $p = .02$, $\eta_p^2 = .06$. These findings suggest that JTC and cognitive ability share variance in predicting confidence and overconfidence, but that cognitive ability is not responsible for the original JTC findings reported above.

Study 3: JTC and Exuberant Theorizing in Learning

In this next study we examined whether JTC behavior also interacted with a psychological mechanism already shown to underlie the beginner’s bubble in overconfidence. That mechanism is overly exuberant theorizing about how to approach the learning task. That is, people quickly construct theories connecting symptoms to health or illness based on very little evidence and experience, soon possessing quite confident theories about how to succeed at the task. What people fail to realize is just how much error and noise is contained in the small sample of data they have, and so their theories are imperfect, at best, and lead to substantial error. With further experience, people become more accurate in these theories, finding the valid lessons in the data and discarding the noise, and their confidence begins to more closely march toward reality (Sanchez & Dunning, 2018).

However, it is likely that people showing high JTC behavior construct theories about the task even more eagerly than their low JTC counterparts, hence developing the greater overconfidence seen in Study 2. Thus, in Study 3, a replication of the zombie diagnosis task, participants at regular intervals reported the theories they were developing about how to diagnose their patients (i.e., they reported what illnesses, if any, each symptom indicated, starting just before beginning the task and then every twelve trials thereafter). Thus, we assessed for each symptom whether people had formed a theory about the outcome it was connected to (versus stated that they did not know) and how confident they were in their inference. Then we aggregated these notions, which we predicted would follow a cubic trend much like the confidence trend already found (see Sanchez & Dunning, 2018). More important, we predicted that high JTC participants would form more extensive and assured theories than their low JTC counterparts—and do so sooner. These richer and firmer theories would feed into the intensified overconfidence the high JTC participants expressed. These theories, so quickly determined, would not necessarily lead them to accuracy.

Method

Participants. One hundred participants were recruited from Amazon's Mechanical Turk crowdsourcing facility. This study was conducted in two parts over two days. Participants received $5 for their participation of part one. For completion of part two, participants received $3 and had the chance to win an additional $3 if they achieved an overall accuracy level of 80%

in the medical diagnosis task described below. Nine participants were excluded from the analysis because they failed to complete part 2. Also, one participant never varied in their diagnostic confidence estimates, suggesting the measure was being ignored, and thus was excluded. The sample consisted of 66% men and 34% women, with 4% wishing to not disclose their gender.

Procedure. In this study we used the same jumping to conclusions measure and zombie task from Study 1, examining how confidence and accuracy waxed and waned over 60 trials of trying to diagnose zombie diseases with feedback after each trial.

Theory Development. However, we added questions to test how quickly participants developed partial to full-blown theories about how medical symptoms connected to possible diagnoses. To do this, we embedded questions at six points throughout the study. At these time points, participants answered 16 questions about their medical theories on diagnosing zombie disease, and how strongly they held these theories. Participants were presented with the eight individual symptoms used in the task, and asked for each whether it indicated a MZD diagnosis, a TS-19 one, both, neither (i.e., the person is healthy), or was irrelevant. If they indicated an answer, they then rated their confidence in that answer from 1(not at all) to 5(certain) that they were right. Finally, participants instead were allowed to answer for each symptom that they did not know.

From these responses, we constructed a scale of theory development. If participants gave an answer, we gave them a score based on their confidence (i.e., a score from 1 to 5). If participants stated they did not know, they received a score for that symptom of 0. We then summed all participant scores across all eight symptoms to indicate the total quantity of theorizing a participant had done up to that point. As such, a person's theory development quantity score could range from 0 (refused to provide any answer for any symptom) to 40 (offered answers for all eight symptoms of which they were completely certain).

Importantly, this overall theory development score could be bifurcated into two components. One part of the score represented theory development for those symptoms in which participants gave a correct answer about the outcome the symptom indicated. The other was for those instances in which the participant gave an erroneous answer. By subtracting the latter score from the former, we could assess the quality of theory a participant had developed.

As such, we probed the quantity and quality of participant theorizing at six different points during the medical diagnosis task. Participants reported their theories just before beginning the diagnosis task, and then probed again after each 12 diagnosis trials, with the last report occurring right after the 60th and final trial.

Results and Discussion

Learning and Performance. We conducted the same analyses as in Study 2 and replicated our findings. When we conducted a mixed model logistic regression including the linear trend of experience and JTC, we found a gradual linear increase in learning as participants performed the probabilistic task, $b = .006$, $se_b = .002$, $p <.001$, $OR = 1.006$ (see left panel of Figure 5). We also found that low jumpers achieved greater accuracy across all of the trials, $b = -.32$, $se_b = .06$, $p < .001$, $OR = .73$.[15] Again, when we tested for an interaction effect between experience and JTC it was nonsignificant, $b = .002$, $se_b = .002$, $p = ns$, suggesting that the rates of learning between high and low jumpers did not differ.

Confidence and Overconfidence. We next repeated the same analyses between jumping to conclusions, experience, and confidence as was done in Study 2. Again, we replicated our

[15] When we added a quadratic component and a JTC X quadratic trend to the analysis, we found again a significant quadratic trend, $b = -.00038$, $se_b = .0001$, $p < .001$, but it did not improve the fit of the overall model, Δ BIC = +22.

main findings. Overall, participants were overconfident, on average reporting an average confidence of 63.1% (*SD* = 14.7) but achieving an accuracy rate of 56.4% (*SD* = 11.0), $t(89) = 3.38$, $p = .001$, $d = .72$. In terms of experience (see Table 6), the linear, quadratic, and cubic time trends were significant, with high jumpers exhibiting significantly greater confidence overall, $b = 5.13$, $se_b = 2.22$, $p = .023$. In terms of JTC X time trend interactions, JTC behavior just missed interacting significantly with the linear time trend, b = -.14, $se_b = .07$, $p = .053$, but did interact with the quadratic trend, $b = .004$, $se_b = .002$, $p = .041$. Figure 5 (right panel) shows how overconfidence developed over time for those showing high (+1 SD) and low (-1 SD) JTC behavior, largely mimicking the pattern found in Study 2.

Further when we then subjected overconfidence scores to a fuller mixed model analysis (random intercept and slope) including JTC as well as the linear, quadratic, and cubic components of experience as fixed variables and participant as a random variable. We again observed a main effect for JTC behavior, $b = 12.91$, $se_b = 2.61$, $p < .001$, $\eta_p^2 = .21$, and an interaction with the linear trend in experience, $b = -.38$, $se_b = .14$, $p = .007$, $\eta_p^2 = .07$, though the interaction effect with the cubic trend was not significant. High jumpers were again more overconfident than low jumpers, but the gap closed by about 10% by the end of the diagnosis task.[16]

Theory Development. We next examined the theories that participants developed about how to approach the diagnosis task, asking how quickly they formed those theories and whether

[16] As vestige left from the survey of Study 2, participants again completed the Ammons intelligence test (*M* = 41.7, *SD* = 5.6, with 3 outliers omitted). In this study, this measure of cognitive ability correlated with accuracy, $r(84) = .44$, $p < .01$, and overconfidence, $r(84) = -.30$, $p < .001$, but not confidence, $r = -.05$. Thus, we re-ran our JTC analyses after controlling for cognitive ability as measured by this test. All significant JTC findings reported in the text remained significant (*p*s < .04), save the JTC X Quadratic trend on confidence. Cognitive ability retained only a direct and negative effect on accuracy in these analyses, $b = .02$, $se_b = .007$, $p = .002$, $OR = 1.02$. That is, with cognitive ability controlled, the relation of JTC to confidence, accuracy, and overconfidence was almost entirely unaffected.

those theories explained the beginner's bubble in overconfidence as well as the overly inflated confidence expressed by those exhibiting high JTC behavior.

Quantity of Theorizing. We first subjected participants' overall quantity of theorizing to a mixed model (random intercept, random slope) analysis looking at the linear, quadratic, and cubic trends of experience, JTC behavior, plus any interaction between. JTC was positively correlated with total amount of theorizing overall, $b = 2.32$, $se_b = 1.10$, $p = .040$, $\eta_p^2 = .05$. All time trends were significant: $b = .59$, $se_b = .13$, $b = -.52$, $se_b = .07$, $b = .29$, $se_b = .04$, for linear, quadratic, and cubic trends, respectively, all ps < .001, all η_p^2s > .19. In addition, JTC interacted with each trend: $b = -.37$, $se_b = .19$, $b = .36$, $se_b = .11$, $b = -.16$, $se_b = .05$, all ps < .05, all η_p^2s > .09. Figure 6 (left panel) shows how the quantity of theory development emerged for those with both high (+1 SD) and low (-1 SD) JTC behavior. Of key import, theory development followed a cubic trend through time, much like confidence. In addition, those exhibiting high JTC behavior expressed more theory overall, and especially before they had any experience whatsoever in the task, $r(87) = .42$, $p < .0001$. That is, high JTC participants jumped to conclusions about how to diagnose zombies in our artificial world even before they had any scrap of data to go on.

What about the quality of those theories? We subjected our measure of theory quality (accurate minus inaccurate theory) to the same mixed model analysis as above. These analyses produced a significant linear, $b = .51$, $se_b = .14$, $p < .001$, $\eta_p^2 = .13$, and quadratic trend over experience, $b = -.40$, $se_b = .13$, $p < .001$, $\eta_p^2 = .18$, as well as a main effect of JTC, $b = -6.18$, $se_b = 1.33$, $p < .001$, $\eta_p^2 = .19$. As seen in Figure 6 (right panel), those with high JTC behavior possessed more inaccurate theories than did their low JTC counterparts.

Did quantity or quality of theory development explain the link between JTC and overconfidence? To assess this question, we conducted a mediational analysis. First, for each participant, we aggregated their judgments to determine their individual level of overconfidence. Then we predicted overconfidence from JTC, b = 12.90, se_b = 2.61, p < .001, η_p^2 = .22. Second, we computed overall theory quantity and quality for each participant. JTC predicted both, b = 2.33, se_b = 1.10, p =.04, η_p^2 = .05, for theory quantity, and b = -6.19, se_b = 1.32, p < .001, η_p^2 = .20, for theory quality. Finally, to determine mediation, we examined whether JTC had any indirect effects on overconfidence that came through theory quantity or quality. Using the *lavaan* package in *R*, bootstrapping 95% confidence intervals via 1000 sampling iterations, we found that an indirect effect through theory quality, *ab* = 7.01, [2.26, 13.12], but not theory quality *ab* = 5.03, [-0.63, 11.68]

In sum, Study 3 replicated the results of Study 2 and explained, at least in part, the differences between high and low jumpers. High jumpers are more overconfident as they learned the task because they developed theories about how to approach that task more quickly than low jumpers—in fact, they especially did so before they received any evidence from experience. In addition, the quality of the theories produced by those with high JTC behavior was significantly lower than those of their low JTC counterparts—not a surprise, given that these theories were developed before adequate evidence of their validity had been collected—which heightened their overconfidence.

Study 4: Intervention in Learning

The first three studies revealed links between JTC behavior and judgment errors and biases. In Study 4, we went a step further, asking if we can debias one of those judgmental biases by borrowing intervention materials used in schizophrenia treatment to inhibit JTC behavior. As such, we modified materials that have been used successfully in schizophrenia treatment to inhibit JTC behavior to see if they also reduce overconfidence in the learning paradigm within a nonclinical population. In essence, can high jumpers be trained not to jump to conclusions, and thus make more accurate or less confident judgments?

Specifically, assessment studies have shown that Metacognitive Training (MCT) interventions briefly improve reasoning with people that are currently suffering from delusions or with schizophrenia spectrum psychosis (Garety et al, 2014; Waller, Freeman, Jolley, Dunn, & Garety, 2011). We borrowed a subset of MCT materials and tested whether completing a mini-intervention emphasizing the importance of not jumping to conclusions would lead to less overconfidence in judgments. While participants completed the lecture, they performed a battery of brainteaser type puzzles that usually lead to judgmental errors and biases. Their errors are pointed out to them and the importance of considering alternative hypotheses is stressed. The next day, participants engaged in the same probabilistic learning task used in Studies 2 and 3, to see if the intervention would reduce the degree of overconfidence participants expressed. We delayed the second part of the experiment a day because we wanted to minimize demand effects in our study. Further, in this manner we could test whether the lecture truly served to change perceptions of their judgments.

Method

Ethics Statement. Study 4 was approved by the Institutional Review Board of Cornell University under the title Metacognitive Training (MCT) to Reduce Judgment Error (Protocol #1503005461).

Participants. In the experimental group, 82 students from Cornell University participated for course credit and $10. They also had the chance to win up to $3 if they achieved certain accuracy levels across all trials. This study was conducted in two parts over two days. Further, we obtained a baseline condition for comparative purposes after the intervention was conducted.

Those participants ($n = 40$), received course credit and had the chance to win the $3. As in the previous study, one participant's confidence judgment was excluded because the participant never varied in their confidence in the probabilistic learning task.

Intervention. All of the materials used throughout all of these tasks were adapted from Moritz and Woodward's (2007) MCT package, with the exception of the video that explained system 1 and system 2 (Minute Videos, 2015). This training program provides an education on jumping to conclusions behavior and belief inflexibility. Furthermore, it teaches a number of techniques and strategies for reducing these biases. This intervention has been used together with researcher or in instructor-led training. We modified the intervention and made it completely computer-based. Materials can be obtained here at the following website: www.uke.de/mct. We should note that we altered those materials so they will not exactly be the materials used. However, you can contact the authors of this article if you would like to see the exact materials used. Participants completed the intervention alone on a computer. To increase engagement in the task video clips were added to the training session.

Task One: Education Program

This was a short education on jumping to conclusions behavior. This portion of the training introduced the idea that everyone jumps to conclusions at some point. Also jumping to conclusions can even be beneficial.

In this portion of the task participants were first instructed that sometimes people make snap judgments too hastily and this causes them to come to wrong conclusions. To illustrate this they were told to watch a video and to try their best to pay close attention. The "Whodunnit" video obtained from the MCT intervention was used (Moritz & Woodward, 2007). The video showed a detective trying to solve a murder similar to the movie clue. After the clip ended, they were asked if they had noticed the 21 changes that had occurred in the scene as they were watching it. As the camera panned, there were numerous changes made to the background. For example, the man that lay on ground dead was replaced by another actor. The video was shown again to the participants but without differing camera angles so they could see the changes to the scenario they hadn't initially detected.

Participants were then educated about what jumping to conclusions behavior is. They were told that everyone at times is vulnerable to this tendency. Further they were shown materials about how jumping to conclusions promotes misinterpretations during psychosis and shown examples. Then they were shown a short video about a doctor diagnosing a patient while family members made snap judgments. Next they were educated about why people make snap judgments and learned about system 1 and system 2 decisions and, watched a short video educating them about jumping to conclusions and how this relates to system 1 and system 2 decisions (Minute Videos, 2015). They were then instructed about the many reasons why people jump to conclusions, the situation in which people are most likely to make hasty decisions, and shown examples. Next they were shown a video of a visual illusion and shown the correct answer to the visual illusion. The purpose of this task was to show participants the value of slowing down their judgments, when making snap decisions people often miss important pieces of information in their environments.

Then they learned about how jumping to conclusions behavior relates to urban legends and conspiracy theories. Lastly, they were taught the consequences of holding these beliefs and how to prevent themselves from jumping to conclusions.

Task Two: What's the Picture?

The purpose of this task was to demonstrate to participants that reaching conclusions can be difficult when they do not have access to complete information. In this task participants saw a series of pictures. Each picture revealed another aspect of the picture. Participants had six options and needed to decide what the picture depicted and how likely it is that each of the six options will be the final picture (Moritz & Woodward, 2007). Then they performed a similar task in which participants looked at paintings and had to guess from four possible choices what the title of the painting was and how confident they were in their judgment. They were given immediate feedback of the correct answer.

Task Three: Optical Illusions

In this task participants saw optical illusions. They were pictures that had multiple interpretations. They wrote down what they saw. Again, they received immediate feedback pointing to all of the details in the pictures that they might have initially missed. This task encourages participants to slow down their judgments.

Task Four: Jumping to Conclusions Ramifications

Participants then watched a short movie clip of a male runner at a park complimenting an attractive female's artwork. She is shown sitting on a park bench looking down sketching in her notepad. While he speaks to her, she does not look up. He becomes irate when she ignores him. Finally, she notices him screaming at her, but he runs off before she can communicate with him. In fact, she is deaf. She leaves him a picture along with an explanation. At the end of the video you see the male runner feeling quite foolish. Participants were instructed that much like in this video they watched, sometimes people misinterpret situations. They were asked to write about when they jumped to a conclusion and they were wrong and to tell us the things they wish they would have considered but didn't in that situation.

Procedure. On day one, participants completed the same jumping to conclusions task used in the previous studies herein before they began the jumping to conclusions intervention. The materials and procedures of the medical diagnosis learning task on day two were identical to those used on day two of Study 2. One participant's confidence measure was excluded from our analysis because confidence measurements never varied.

Results and Discussion

Experience, Accuracy, and Confidence. Did the intervention quell overconfidence? Did JTC behavior continue to have an influence after the intervention? To address these questions, we first subjected accuracy scores to a mixed-model analysis (random intercept, random slope) that included the intervention, JTC, and the linear component of experience. Results are depicted in the left panel of Figure 7. As seen in the panel, we observed a linear increase in accuracy over time, b = .006, se_b = .0015, p < .001, OR = 1.01. The intervention did not have a significant impact, b = -.028, se_b = .056, *ns*, OR = .97, but JTC did remain significantly related to accuracy in diagnosis, b = -.103, se_b = .04, p = .009, OR = .90. We also tested for an intervention X JTC interaction, but it was not statistically significant. Thus, our intervention did not impact objective performance.

Results of a similar analysis on confidence, however, produced differing results. First, all three time trends were significant (see Figure 7, left panel, and Table 6). When we predicted confidence from the intervention and JTC, the intervention significantly lowered confidence, b = -5.18, se_b = 2.35, p = .03, with JTC no longer related to confidence, b = 2.88, se_b = 3.49, *ns.* No interaction between JTC and intervention arose, b = .66, se_b = 3.49, *ns*.

However, in terms of overconfidence, a similar analysis revealed effects for both the intervention and JTC (see Figure 7, right panel). The intervention lowered it, b = -6.39, se_b = 2.97, p = .03, η_p^2 = .04, but JTC remained linked to it, b = 5.54, se_b = 2.10, p = .009, η_p^2 = .06.

There was no significant JTC by intervention interaction. This finding is consistent with schizophrenia research, in that no research exists to our knowledge indicates that MCT only works on those with showing high JTC behavior.

Figure 8 depicts the role played by the intervention on confidence and accuracy independent of time trends. It shows how the intervention reduced confidence but had little impact on accuracy—with JTC retaining its relationship with accuracy. The impact of the intervention on confidence allowed it also to reduce overconfidence, but JTC retained its relationship to overconfidence through its continued correlation with accuracy.

Theory Development. We next looked at the role played by the intervention, JTC, and experience on the quantity and quality of theory development (see Figure 9). A mixed model analysis with the intervention, JTC, time trends, and interactions between revealed that the quantity of theory development followed the largely cubic trend found previously. All time trends were significant: linear, $b = 1.11$, $se_b = .09$, $p < .001$, $\eta_p^2 = .58$; quadratic, $b = -.64$, $se_b = .07$, $p < .001$, $\eta_p^2 = .43$; and cubic, $b = .30$, $se_b = .04$, $p < .001$, $\eta_p^2 = .34$, and the intervention reduced it overall, $b = -2.98$, $se_b = .97$, $p < .003$, $\eta_p^2 = .08$ (see Figure 9, left panel). JTC was no longer related to quantity of theory development, except for an interaction with the linear time trend, $b = -.35$, $se_b = .14$, $p =.02$, $\eta_p^2 = .05$, which suggested that low JTC behavior developed more theory as the diagnosis task continued (not a surprise, given the high JTC habit of already forming a lot of theory before they started the task).

In terms of theory quality (see Figure 9, left panel), a similar analysis produced different results. The linear, $b = .88$, $se_b = .12$, $p <.001$, $\eta_p^2 = .33$, and quadratic, $b = -.33$, $se_b = .08$, $p <.001$, $\eta_p^2 = .13$, time trends were significant. The intervention, however, failed to improve the quality of theoretical development. Here, it was JTC that mattered, in that those showing high

JTC behavior reported theories that were worse than the low JTC counterparts, b = -4.18, se_b = 1.10, p <.001, η_p^2 = .12. The only interaction that rose to significance was a JTC X linear trend one, b = -.51, se_b = .18, p = .005, η_p^2 = .07, in which the theoretical quality of low JTC participants rose throughout the diagnosis task faster than it did for high JTC peers.

In sum, the intervention reduced theorizing overall (quantity), but it did nothing to improve quality. JTC was the variable that governed quality, although its relationship to quantity was not significant.

Finally, we asked whether reducing theoretical quantity was the reason why participants were less overconfident in the intervention condition? We have already shown above that (a) the intervention reduced overconfidence, and that (b) the intervention also reduced the supposed mediator, theoretical quantity. The final question to ask is whether the intervention had an indirect effect on overconfidence via theoretical quantity. To address this, we conducted a mediational analysis using the *lavaan* package in R, bootstrapping 95% confidence intervals via 1000 iterations. This analysis revealed a successful mediation: a significant indirect effect flowing through theoretical quantity, *ab* = -3.22 [-5.92, -0.99]. Theoretical quality, in comparison, failed as a mediator, *ab* = -0.51 [-0.68, 2.02].

Summary. Taken together, these findings suggest our intervention was a success, albeit qualified. Prompting participants to avoid JTC behavior lowered their confidence and their overconfidence. Thus, there is a causal link between JTC behavior and confidence in judgment. Intervening to reduce the behavior affected the tendency toward high confidence among people from a nonclinical population. However, the intervention was more blunt than surgical, in that it lowered confidence of everyone, not just that of high jumpers. It also prompted participants, in the end, to be generally underconfident rather than more sensitive to when their judgments were right versus wrong. In the end, JTC was still linked to overconfidence through the link between JTC behavior and accuracy, which the intervention did nothing to ameliorate.

General Discussion

Life is the art, as once put by British novelist Samuel Butler (1912), of drawing sufficient conclusions from insufficient premises. People must reach judgments and make predictions under the shadow of uncertainty, and psychological research over the past half century has observed and commented on their ability to do so. People do at times show remarkable competence in their facility to reach decisions when faced with incomplete or imperfect information, but their conclusions are also often marked by orderly error, false belief, and unwarranted confidence (Gilovich, Griffin, & Kahneman, 2002; Nisbett & Ross, 1980;).

In this manuscript, we asked whether there were any systematic variation in a nonclinical population that could indicate who was vulnerable to judgmental deficits. Our starting point was the literature on schizophrenia, which has long pointed to a behavior or tendency associated with being delusion-prone. That tendency is jumping to conclusions, in which individuals leap to conclusions based on only the barest of data (Dudley et al., 2015; Garety & Freeman, 1999; Huq et al., 1988; Moritz & Woodward, 2005). Schizophrenia patients who display this behavior are more prone to endorsing delusions (Garety et al., 2005), experiencing hallucinations and perceptual anomalies (Freeman, Pugh, & Garety, 2008), expressing paranoid beliefs (Freeman, Pugh, & Garety, 2008; Garety et al., 2015; Moulding et al., 2016), and reaching faulty inferences (Moritz &Woodward, 2006a; Moritz,Woodward, & Rodriguez-Raecke, 2006) than counterparts who wait for more information before making decisions in probabilistic judgment.

Throughout all of our studies, we have treated JTC as a behavior in a particular task, with present research shedding light on the processes that underlie JTC in that situation. Across five studies, we assessed people's tendency to jump to conclusions in a JTC task among a nonclinical sample and then conducted an extensive survey of its judgmental correlates within that sample. We found that people from a sample of the general public who displayed JTC behavior were more likely to commit classic judgmental errors suggesting interference by the quick and intuitive operation of "system 1" reasoning. More specifically, in Study 1, they were more likely to commit belief bias in logical reasoning (Evans, Barston, & Pollard, 1983; Evans & Curtis-Holmes, 2005), to display denominator neglect in gambling (Denes-Raj & Epstein, 1994; Reyna & Brainerd, 2008), and to make a greater number of errors on the Cognitive Reflection Test (Frederick, 2005; Toplak et al., 2014). They endorsed unlikely conspiracy theories as true, including theories about doctors and medicine that are associated with avoiding conventional medical care (Oliver & Wood, 2014). They also reported a higher degree of belief in the paranormal. In addition, they were more overconfident in their answers to a civics quiz than their low JTC peers.

Studies 2 and 3 examined the implications of JTC behavior for a key task in everyday life: probabilistic learning. Participants gained experience and feedback as they completed sixty trials of a medical diagnosis task. All participants learned to more accurately diagnose patients with more experience, but high JTC participants were less accurate overall than their low JTC counterparts. Despite this reduced accuracy, they were more confident in their diagnoses, leading them to be significantly more overconfident in their performance on the task. It appears, in Study 3, that high jumpers raced prematurely to strategies about how to diagnose—strategies that were filled with noise and error—relative to low jumpers. Because of this, although they improved in their diagnoses at the same rate as their low jumping peers, they started from an accuracy deficit from which they never fully recovered. Their theories about how to approach the task remained much more contaminated with misleading notions than that of their low jumping peers.

Finally, in Study 4, we adapted for use with the general public an intervention designed to quell JTC reasoning among schizophrenia patients, focusing on Metacognitive Training (MCT) (Moritz & Woodward, 2007). The intervention prompted participants away from overconfidence as they completed the medical diagnosis task. Indeed, MCT materials were successful in making participants actually underconfident, rather than overconfident, in their success. However, these materials had the same impact on all participants, not just high jumpers, and so the intervention did not remove differences in confidence between those with high and low JTC behavior.

Taken together, these findings suggest that JTC behavior, a tendency put under close scrutiny for over twenty years in schizophrenia research, may be a variable that has wide-ranging cognitive and social correlates within a nonclinical population. A recipe of increased error and greater false belief coupled with high confidence suggests a behavioral tendency that might predispose people toward “knowledge corruption” (Moritz & Woodward, 2002), as it is known in schizophrenia research, or miscalibrated overconfidence, as it is known in the judgment and decision making world (Lichtenstein et al., 1977). Further analyses showed that links we found between JTC behavior and judgmental deficits were largely independent of other individual differences, such as need for closure, need for cognition, and general cognitive ability.

Mechanism: Automatic versus Controlled Thinking

In further analysis from Study 1, we explored whether the link between JTC and judgmental deficits came from an overreliance on “system 1” or automatic processing (as suggested by the hypersalience hypothesis) or on a lack of “system 2” or controlled processing (as suggested by a “liberal decision criterion” account). We took participants’ performance on the belief bias and denominator neglect tasks and decomposed them into their automatic and controlled components. We discovered that high JTC behavior was primarily associated with less controlled processing.

More informative, when we conducted mediational analyses, we discovered that lack of controlled processing consistently mediated the link between JTC and judgmental shortcomings. Level of automatic processing failed to produce effects that were consistent across both studies. In short, it appears a connection between automatic processing and JTC is nonexistent or at best undocumented. High JTC participants, it appeared, did not engage in controlled processing that can countermand the impact of that initial automatic processing.

This is not to say that the minds of those with high JTC behavior are inert, impulsive, or unmotivated. In Study 3, they showed just as much evidence of thought and calculation as their peers, at least in terms of the theories they reported they were using to tackle the medical diagnosis task. Their theories were not sparse or impoverished, but instead were at least as elaborate and complete as their low JTC counterparts. Indeed, as an exploration, we tested just how frequently participants changed their theories (e.g., changed their minds about what a specific symptom indicated) across the 60 trials in Study 3, and found that the frequency of these revisions failed to correlate with JTC, $r(88) = .14$, *ns.*

Thus, if anything distinguished high JTC thinking, it was the willingness to form these elaborate theories with undue haste, often even before the task began and ahead of any feedback or experience to draw on. In fact, before seeing their first patient, the degree to which participants already had ideas about which symptoms diagnosed zombie diseases was correlated .42, $df = 88$, $p < .001$, with JTC behavior. In short, at the beginning of the task, those with high JTC behavior were not sluggish in their theorizing, they were overly exuberant to get going—

and then simply too quick to conclude. Of course, with no valuable experience or feedback to guide them, much of what they surmised was wrong, and measures of theory quality suggested they never recovered from those initial hasty and misguided inferences.

Further, data from across all five studies also suggest that JTC effects were not explained by differences in motivation or impulsivity. First, scales designed to measure motivations (need for cognition, need for closure) did not explain any of the relationship between JTC and judgmental outcomes. Second, removing participants more frequently in Study 1b who explicitly failed attention checks, thus showing low motivation and care, did nothing to diminish differences we found between high and low jumpers relative to Study 1a, where attention checks were not used. Third, high jumpers showed just as much evidence of motivation or care for tasks as low jumpers. They asked to see more fish when the task required more evidence (in the 65%/35% version of the lake) than when it was easier (in the 80%/20% version of the lake), just like low jumpers (Studies 1a and 1b). Taken together, all these data suggest that high jumpers were paying attention and motivated to do just as well on the tasks as were low jumpers.

Similarly, in the medical diagnosis task (Studies 2 – 4), with experience, high jumpers improved at their diagnostic accuracy at the same rate as low jumpers; they just required less evidence to reach a decision (Studies 1a or 1b). In the medical diagnosis task, it appears that high jumpers achieved less overall accuracy not because of their rate of learning but because they sprung into premature theorizing about the task when they had no or little data to base that theorizing on. That premature jump to theory produced a number of misguided notions that appeared to handicap their performance as they progressed through the task (see Figure 6), preventing them from catching up to their low jumping peers.

Implications for Knowledge Corruption (The Dunning-Kruger Effect)

Given the range of judgmental flaws we attached to JTC behavior in this research, it is likely to have implications elsewhere. One specific phenomenon it speaks to is the literature on the Dunning-Kruger effect (Dunning, 2011; Dunning et al., 2003; Kruger & Dunning, 1999). As understood in popular culture, the effect refers to the observation that incompetent individuals often fail to realize just how poorly they are performing. They make many mistakes but judge their performances almost as favorably as top performers judge theirs.

One open question about the effect has been whether it is local or general in nature. The original account of this effect emphasized its local nature. People all have some circumscribed pockets of incompetence they fail to recognize. Thus, everyone, sooner or later, is vulnerable to the effect (Kruger & Dunning, 1999), depending on where their specific intellectual and social deficiencies lay. The original account of the effect, however, left open the possibility that there might be some cognitive or personality factors that predisposed some segment of the nonclinical population toward the Dunning-Kruger effect in a more general way.

The work herein suggests that an inclination toward high JTC behavior might be one such difference that predisposes some people toward the Dunning-Kruger effect. Specifically, across all the studies, high JTC participants expressed levels of confidence suggesting they had little idea of just how many errors they were making. They were the ones most likely to be unskilled and unaware of it.

Connections Between Clinical and Social Psychology

This research also suggests proposals for further crosstalk between research on psychopathology and work being conducted in social, personality, and cognitive psychology within the general population. Could potentially stimulating opportunities exist simply by these fields paying somewhat more attention to each other? For example, schizophrenia researchers have long shown that JTC behavior is associated with delusion, psychosis, and impaired reasoning for schizophrenia patients. In this manuscript, we noted that observation and asked if JTC behavior is similarly associated with the types of judgmental errors and biases so often documented in the social and cognitive literatures. We found a broad suite of implications for JTC behavior among the nonclinical populations.

Are there other insights lying within clinical psychology that could help explain other issues in social and personality psychology? One need not look far to find other example of a potentially telling tendency in the literature on schizophrenia. Namely, proneness to delusion is also linked to an independent tendency known as the bias against disconfirmatory evidence (BADE). Specifically, when reading a story that makes one interpretation or conclusion likely, delusion-prone schizophrenia patients fail to revise their interpretations in light of information suggesting a completely different conclusion. Like JTC behavior, BADE has been connected to delusion proneness in both clinical and non-clinical samples (Woodward, Buchy, Moritz, & Liotti, 2007). Could it also be similarly associated with knowledge corruption, unrealistic beliefs, and other judgmental biases found in non-clinical populations?

In addition, a potential crosstalk between findings in clinical psychology and social/personality psychology, as well as cognitive psychology, might prove profitable in both directions. Work on psychopathology can suggest psychological phenomena or dynamics to explore among the general population, but work in the general population may prove valuable toward understanding the specific psychology in play with these phenomena. That is, work on schizophrenia, for example, is hampered in that the condition affects only one percent of the population (Cooper et al., 2007). Thus, doing research completely within only this sparse population can slow the research process in its pursuits of explanatory insights.

Work within the general population might be able to provide clearer, more precise, and faster answers about the cognitive, neurological, and behavioral mechanisms associated with JTC that may prove valuable to researchers and practitioners working with special populations in the clinical realm as well as with the general population.

However, perhaps deeper than this, future work with non-clinical samples can assess the extent to which jumping to conclusions represents a broad trait or just a behavior specific to tasks like the fish-lake task we presented here, which is very much like the task used in schizophrenia research (Dudley et al., 2015). Performance on the task does show some signs of being a trait. Past work has shown it has remarkable temporal stability (So et al., 2012), and the data presented here shows that it correlates with a broad suite of judgment tasks explored within psychology. However, the question of whether JTC exhibits cross-situational generality remains an intriguing research question. For example, a recent industry survey showed that 41% of Americans visit only one dealership before buying a car—a rather consequential decision (Cox Automotive, 2019). Does that behavior bear any relation to behavior on the task we presented participants in these studies, or to any type of decisions that people might confront?

In sum, one wonders if some closer relationship between the two fields might be profitable. This is not a suggestion to merge journals or fields; far from it. Instead, it is a modest suggestion that the two fields pay more attention to developments in the other, and develop potentially valuable areas of discussion. Would it worth it to do so? We jump to no conclusions, but suggest that more exploration of the possibility may be worthwhile.

References

Aghotor, J., Pfueller, U., Moritz, S., Weisbrod, M., & Roesch-Ely, D. (2010). Metacognitive training for patients with schizophrenia (MCT): feasibility and preliminary evidence for its efficacy. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 207-211.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Ammons, R. B., & Ammons, C. H. (1962). The Quick Test (QT): Provisional manual. *Psychological Reports, 11,* 111-161.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*, 258–290.

Balzan, R. P. (2016). Overconfidence in psychosis: The foundation of delusional conviction?. *Cogent Psychology*, *3*, 1-11.

Balzan, R. P., Delfabbro, P. H., Galletly, C. A., & Woodward, T. S. (2014). Metacognitive training for patients with schizophrenia: Preliminary evidence for a targeted, single-module programme. *Australian & New Zealand Journal of Psychiatry*, *48,* 1126-1136.

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

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Explaining delusions: A cognitive perspective. *Trends in Cognitive Sciences*, *10*, 219-226.

Benjamin, D. J., Rabin, M., & Raymond, C. (2016). A model of nonbelief in the law of large numbers. *Journal of the European Economic Association, 14*, 515–544.

Bentall, R. P. (1992). A proposal to classify happiness as a psychiatric disorder. *Journal of Medical Ethics, 18*(2), 94-98.

Bosson, J. K., Swann Jr, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited?. *Journal of Personality and Social Psychology*, *79*, 631-643.

Broome, M. R., Johns, L. C., Valli, I., Woolley, J. B., Tabraham, P., Brett, C., Valmaggia, L., Peters, E., Garety, P.A., & Mcguire, P. K. (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *The British Journal of Psychiatry, 191*, s38-s42.

Buchy, L., Woodward, T., & Liotti, M. (2007). A cognitive bias against disconfirmatory evidence (BADE) is associated with schizotypy. *Schizophrenia Research, 90*, 334-337. 012.

Butler, S. (1912). *Notebooks*, ed. Geoffrey Keynes and Brian Hill. London: Jonathan Cape.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment,48*, 306-307.

Camerer, C. F. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *The American Economic Review, 77*, 981-997.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Cooper, S. A., Smiley, E., Morrison, J., Allan, L., Williamson, A., Finlayson, J., Jackson, A., & Mantry, D. (2007). Psychosis and adults with intellectual disabilities. *Social Psychiatry and Psychiatric Epidemiology, 42*, 530-536.

Cox Automotive (2019). *Car buyer journey 2019.* Atlanta: Cox Automotive.

Cull, J. G., & Colvin, C. R. (1970). Correlation between the quick test (Qt) and the WAIS verbal scale in the rehabilitation setting. *Psychological Reports, 27*, 105-106.

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin, 21*, 1083-1092.

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2015). Psychosis, delusions and the "jumping to conclusions" reasoning bias: A systematic review and meta-analysis. *Schizophrenia Bulletin, 42*, 652-665.

Dudley, R. E., John, C. H., Young, A. W., & Over, D. E. (1997). Normal and abnormal reasoning in people with delusions. *British Journal of Clinical Psychology, 36*, 243-258.

Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (vol. 44, pp. 247-296). New York: Elsevier.

Dunning, D., Griffin, D. W., Milojkovic, J. H., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58*, 568-592.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12,* 83-86.

Evans, S., Averbeck, B., & Furl, N. (2015). Jumping to conclusions in schizophrenia. *Neuropsychiatric Disease and Treatment, 11,* 1615-1624.

Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*, 295-306.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking and Reasoning, 11*, 382–389.

Everett, J., Lavoie, K., Gagnon, J. F., & Gosselin, N. (2001). Performance of patients with schizophrenia on the Wisconsin Card Sorting Test (WCST). *Journal of Psychiatry and Neuroscience*, *26*, 123–130.

Favrod, J., Rexhaj, S., Bardy, S., Ferrari, P., Hayoz, C., Moritz, S., Conus, P., & Bonsack, C. (2014). Sustained antipsychotic effect of metacognitive training in psychosis: a randomized-controlled study. *European Psychiatry*, *29*, 275-281.

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of JTC behavior in delusions. *Cognitive Neuropsychiatry, 12*, 46-77.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. *Clinical Psychology Review*, *27*, 425-457.

Freeman, D., Pugh, K., & Garety, P. (2008). Jumping to conclusions and paranoid ideation in the general population. *Schizophrenia Research,102*, 254-260.

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of the evidence. *British Journal of Clinical Psychology, 38,* 113-154.

Garety, P. (1991). Reasoning and delusions. *The British Journal of Psychiatry*.

Garety, P. A., & Freeman, D. (2013). The past and future of delusions research: From the inexplicable to the treatable. *The British Journal of Psychiatry, 203*, 327-333.

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., . . . Dudley, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of Abnormal Psychology, 114*, 373-384.

Garety, P., Joyce, E., Jolley, S., Emsley, R., Waller, H., Kuipers, E., . . . Freeman, D. (2013b).

Neuropsychological functioning and jumping to conclusions in delusions. *Schizophrenia Research, 150*, 570-574.

Garety, P., Waller, H., Emsley, R., Jolley, S., Kuipers, E., Bebbington, P., . . . Freeman, D. (2014). Cognitive mechanisms of change in delusions: An experimental investigation targeting reasoning to effect change in paranoia. *Schizophrenia Bulletin, 41*, 400-410.

Gawęda, Ł., Krężołek, Gawęda, M., Olbryś, J., Turska, A., & Kokoszka, A. (2015). Decreasing self-reported cognitive biases and increasing clinical insight through meta-cognitive training in patients with chronic schizophrenia. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 98-104.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology, 18,* 186 –221.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32,* 385–410.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57,* 243–259.

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgments in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology, 40*, 801-812.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30*, 513-541.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. (1968). Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology, 10,* 317–340.

Joyce, D. W., Averbeck, B. B., Frith, C. D., & Shergill, S. S. (2013). Examining belief and confidence in schizophrenia. *Psychological Medicine*, *43*, 2327-2338.

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

Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior, 1967: Aversive stimulation* (pp. 9–31). Coral Gables, FL: University of Miami Press.

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality, 18,* 431–439.

Köther, U., Veckenstedt, R., Vitzthum, F., Roesch-Ely, D., Pfueller, U., Scheu, F., & Moritz, S. (2012). “Don't give me that look”—Overconfidence in false mental state perception in schizophrenia. *Psychiatry Research*, *196,* 1-8.

Kruger, J. M., & 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*, 1121-1134.

Kruglanski, A. W., Atash, M. N., De Grada, E., Mannetti, L., & Pierro, A. (2013) . Need for Closure Scale (NFC) . Measurement instrument database for the social science. Retrieved from www.midss.ie

Langdon, R., Ward, P. B., & Coltheart, M. (2008). Reasoning Anomalies Associated With Delusions in Schizophrenia. *Schizophrenia Bulletin,36*, 321-330.

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation an worldviews in predicting rejection of science. *PLoS One, 10,* e0134773. https://doi.org/10.1371/journal.pone.0134773

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764-766.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In *Decision making and change in human affairs* (pp. 275-324). Springer, Dordrecht.

Minute Videos (2015). *Jumping to conclusions - Thinking fast and slow S01E08*. Retrieved July 11, 2017, from https://www.youtube.com/watch?v=BMPiR2JP2p8

Moritz, S., Andreou, C., Schneider, B. C., Wittekind, C. E., Menon, M., Balzan, R. P., & Woodward, T. S. (2014). Sowing the seeds of doubt: A narrative review on metacognitive training in schizophrenia. *Clinical Psychology Review*, *34*, 358-366.

Moritz, S., Ramdani, N., Klass, H., Andreou, C., Jungclaussen, D., Eifler, S., ... & Zink, M. (2014). Overconfidence in incorrect perceptual judgments in patients with schizophrenia. *Schizophrenia Research: Cognition*, *1*, 165-170.

Moritz, S., Kerstan, A., Veckenstedt, R., Randjbar, S., Vitzthum, F., Schmidt, C., Heise, M., & Woodward, T. S. (2011). Further evidence for the efficacy of a metacognitive group training in schizophrenia. *Behaviour Research and Therapy*, *49*, 151-157.

Moritz, S., & Van Quaquebeke, N. (2014). Are you sure? Delusion conviction moderates the behavioural and emotional consequences of paranoid ideas. *Cognitive Neuropsychiatry*, *19*, 164-180.

Moritz, S., Veckenstedt, R., Andreou, C., Bohn, F., Hottenrott, B., Leighton, L., Köther, U., Woodward, T.S., Treszl, A., Menon, M., & Schneider, B. C. (2014). Sustained and "sleeper" effects of group metacognitive training for schizophrenia: A randomized clinical trial. *JAMA Psychiatry*, *71*, 1103-1111.

Moritz, S., Veckenstedt, R., Bohn, F., Hottenrott, B., Scheu, F., Randjbar, S., J., Köther, U., Woodward, T.S., Treszl, A., & Andreou, C. (2013). Complementary group Metacognitive Training (MCT) reduces delusional ideation in schizophrenia. *Schizophrenia Research*, *151*, 61-69.

Moritz, S., Veckenstedt, R., Randjbar, S., Vitzthum, F., & Woodward, T. S. (2011). Antipsychotic treatment beyond antipsychotics: Metacognitive intervention for schizophrenia patients improves delusional symptoms. *Psychological Medicine*, *41*, 1823-1832.

Moritz, S., & Woodward, T. S. (2002). Memory confidence and false memories in schizophrenia. *The Journal of Nervous and Mental Disease, 190,* 641-643.

Moritz, S., & Woodward, T. S. (2004). Plausibility judgment in schizophrenic patients: Evidence for a liberal acceptance bias. *German Journal of Psychiatry, 7,* 66-74.

Moritz, S., & Woodward T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology 44*, 193–207.

Moritz, S., & Woodward, T. S. (2006a). The contribution of metamemory deficits to schizophrenia. *Journal of Abnormal Psychology, 115*, 15-25.

Moritz, S., & Woodward, T. S. (2006b). Metacognitive control over false memories: A key determinant of delusional thinking. *Current Psychiatry Reports*, *8*, 184-190.

Moritz, S., & Woodward, T. S. (2006c). A generalized bias against disconfirmatory evidence in schizophrenia. *Psychiatry Research*, *142*, 157-165.

Moritz, S., & Woodward, T. S. (2007). Metacognitive training in schizophrenia: From basic research to knowledge translation and intervention. *Current Opinion in Psychiatry, 20*, 619-625.

Moritz, S., Woodward, T. S., & Rodriguez-Raecke, R. (2006). Patients with schizophrenia do not produce more false memories than controls but are more confident in them. *Psychological Medicine*, *36,* 659-667.

Moulding, R., Nix-Carnell, S., Schnabel, A., Nedeljkovic, M., Burnside, E. E., Lentini, A. F., & Mehzabin, N. (2016). Better the devil you know than a world you don't? Intolerance of uncertainty and worldview explanations for belief in conspiracy theories. *Personality and Individual Differences*, *98*, 345-354.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Oliver, J. E., & Wood, T. (2014). Medical conspiracy theories and health behaviors in the United States. *JAMA Internal Medicine, 174*, 817-818.

Ono, K. (1987). Superstitious behavior in humans. *Journal of the Experimental Analysis of Behavior, 47,* 261–271.

Pantelis, C., Barber, F. Z., Barnes, T. R., Nelson, H. E., Owen, A. M., & Robbins, T. W. (1999). Comparison of set-shifting ability in patients with chronic schizophrenia and frontal lobe damage. *Schizophrenia Research, 37*, 251-270.

Peters, E. R., Joseph, S. A., & Garety, P. A. (1999). The measurement of delusional ideation in the normal population: Introducing the PDI (Peters et al. Delusions Inventory). *Schizophrenia Bulletin, 25*, 553–576.

Petty, R. E., Cacioppo, J. T., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306-307.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review, 46,* 657– 685.

Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders, 9*, 346-355.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89-107.

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33*, 1492-1502.

Ross, R. M., Mckay, R., Coltheart, M., & Langdon, R. (2015). Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data-gathering. *Schizophrenia Bulletin, 41*, 1183-1191.

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology, 114,* 10-28.

Sanford, N., Veckenstedt, R., Moritz, S., Balzan, R. P., & Woodward, T. S. (2014). Impaired integration of disambiguating evidence in delusional schizophrenia patients. *Psychological Medicine*, *44*, 2729-2738.

Savulich, G., Shergill, S., & Yiend, J. (2012). Biased cognition in psychosis. *Journal of Experimental Psychopathology*.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119,* 3.

So, S. H., Chan, A. P., Chong, C. S., Wong, M. H., Lo, W. T., Chung, D. W., & Chan, S. S.

(2015). Metacognitive training for delusions (MCTd): Effectiveness on data-gathering and belief flexibility in a Chinese sample. *Frontiers in Psychology, 6*.

So, S. H., Freeman, D., Dunn, G., Kapur, S., Kuipers, E., Bebbington, P., ... & Garety, P. A. (2012). Jumping to conclusions, a lack of belief flexibility and delusional conviction in psychosis: A longitudinal investigation of the structure, frequency, and relatedness of reasoning biases. *Journal of Abnormal Psychology*, *121*, 129.

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience: JPN*, *35,* 7.

Tobacyk, J. J. (2004). A revised paranormal belief scale. *International Journal of Transpersonal Studies*, *23*(1), 11.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*, 147-168.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.

Van Der Gaag, M. (2006). A neuropsychiatric model of biological and psychological processes in the remission of delusions and auditory hallucinations. *Schizophrenia Bulletin*, *32,* S113-S122.

Ventura, J., Wood, R. C., Jimenez, A. M., & Hellemann, G. S. (2013). Neurocognition and symptoms identify links between facial recognition and emotion processing in schizophrenia: Meta-analytic findings. *Schizophrenia Research*, *151*, 78-84.

Waller, H., Freeman, D., Jolley, S., Dunn, G., & Garety, P. (2011). Targeting reasoning biases in delusions: A pilot study of the Maudsley Review Training Programme for individuals with persistent, high conviction delusions. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 414-421.

Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., & Haudenschield, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behaviour Research and Therapy*, *45*, 1255-1269.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049.

Weinberger, J., & McClelland, D. C. (1990). Cognitive versus traditional motivational models: irreconcilable or complementary? In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior*, Vol. 2, pp. 562-597). New York, NY, US: Guilford Press.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*, 1006–1014.

Winter, D. G. (1999). Linking personality and "scientific" psychology: The development of empirically derived Thematic Apperception Test measures. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 107-124). Washington, DC: American Psychological Association.

Woodward, T. S., Buchy, L., Moritz, S., & Liotti, M. (2007). A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. *Schizophrenia Bulletin*, *33,* 1023-1028.

Woodward, T.S., Moritz, S., Cutler, C., Whitman, J.C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. *Journal of Clinical and Experimental Neuropsychology 28,* 605–617.

Yarritu, I., Matute, H., & Luque, D. (2015). The dark side of cognitive illusions: When an illusory belief interferes with the acquisition of evidence-based knowledge. *British Journal of Psychology, 106,* 597– 608.

Zawadzki, J. A. (2012). Cognitive factors associated with subclinical delusional ideation in the general population. *Psychiatry Research,* 345-349.

Table 1. Simple correlations between measures and tasks used in Study 1a and 1b.

	Measure															
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. JTC		-.14*	-.01	.04	-.33***	.31***	.31***	.15**	.20***	.22***	.13*	.22***	-.11^	-.35***	.32***	-.16**
2. Schizotypy	.08		-.15*	-.07	-.04	.05	-.03	.02	.07	.08	.22***	.14*	-.12*	-.07	-.03	.10^
3. Need for Cognition	-.14**	-.03		-.16**	.13*	-.15**	-.05	-.10^	-.05	-.09	-.02	-.07	.13*	.22***	-.15*	-.08
4. Need for Closure	-.05	.01	-.37***		.07	.01	.09	.14*	.08	.12^	-.03	.07	.08	-.01	.08	-.11^
5. Cognitive Ability	-.43***	-.12*	.11*	.01		-.39***	-.33***	-.27***	-.30***	-.29***	-.15*	-.29***	.33***	.56***	-.39***	.15*
6. CRT	.43***	.03	-.26***	.11^	-.45***		.50***	.24***	.24***	.30***	.25***	.29***	-.25***	-.45***	.31***	-.08
7. Belief Bias	.29***	.01	-.20***	.15*	-.22***	.41***		.28***	.24***	.27***	.25***	.30***	-.16**	-.40***	.33***	-.17**
8. Denominator Neglect	.30***	.02	-.10^	.04	-.36***	.40***	.38***		.18**	.24***	.18**	.24***	-.19**	-.29***	.18**	-.03
9. Medical Myths	.29***	.29***	-.18***	.05	-.38***	.32***	.35***	.37***		.82***	.46***	.91***	-.25***	-.39***	.25***	-.11^
10. Conspiracy Theories	.27***	.29***	-.17***	.02	-.40***	.28***	.30***	.31***	.87***		.42***	.90***	-.30***	-.45***	.27***	-.10^
11. Paranormal Beliefs	.30***	.32***	-.14*	-.01	-.41***	.29***	.33***	.37***	.70***	.67***		.72***	-.21***	-.33***	.21***	-.05
12. Aggregate Oddball Beliefs	.32***	.33***	-.18***	.02	-.43***	.33***	.32***	.39***	.94***	.93***	.86***		-.30***	-.47***	.29***	-.10^
13. Confidence	-.06	-.02	.18***	.02	.24***	-.19***	-.07	-.06	-.14*	-.09	-.09	-.11^		.56***	.21***	-.23***
14. Accuracy	-.36***	-.16**	.18***	-.09	.56***	-.52***	-.31***	-.36***	-.46***	-.41***	-.43***	-.48***	.42***		-.69***	.07
15. Overconfidence	.33**	.15*	-.06	.11^	-.40***	.40***	.27***	.33***	.38***	.36***	.39***	.41***	.32**	-.72***		-.28***
16. Discrimination	-.16**	.04	-.01	.00	.12*	-.08	-.18***	-.14*	-.10^	-.11^	-.14*	-.13*	-.32***	-.05	-.19***	

Note: Correlations from Study 1a are displayed above the diagonal; those from Study 1b are displayed below.

All tests two-tailed: ^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .005$

Table 2. Original correlations between JTC and outcome measures along with correlations after controlling for related constructs (Study 1).

		Controlling For				
Measure	JTC	Schizo-typy	N Cognition	N Closure	CRT	Cognitive Ability
System 1 Biases						
CRT						
Study 1a	.31***	.34***	.34***	.34***		.25***
Study 1b	.43***	.43***	.41***	.44***		.29***
Belief Bias						
Study 1a	.31***	.34***	.34***	.34***	.22***	.27***
Study 1b	.29***	.29***	.27***	.30***	.14*	.22***
Denominator Neglect						
Study 1a	.15**	.17**	.16*	.15*	.08	.09
Study 1b	.30***	.30***	.29***	.30***	.15*	.13*
Oddball Beliefs						
Medical Myths						
Study 1a	.20***	.21***	.20***	.20***	.14*	.11^
Study 1b	.29***	.28***	.27***	.29***	.18***	.15*
Conspiracy Theories						
Study 1a	.22***	.23***	.22***	.22***	.15*	.13*
Study 1b	.27***	.26***	.26***	.28***	.18***	.12*
Paranormal Beliefs						
Study 1a	.13*	.18**	.14*	.14*	.03	.10
Study 1b	.30***	.29***	.28***	.30***	.20***	.15*
Aggregate						
Study 1a	.22***	.25***	.34***	.22***	.13*	.14*
Study 1b	.32***	.31***	.30***	.32***	.21***	.16*
Knowledge Corruption						
Confidence						
Study 1a	-.11^	-.12*	-.13*	-.11	-.04	-.01
Study 1b	-.05	-.05	-.03	-.05	.03	.06
Accuracy						
Study 1a	-.35***	-.35***	-.34***	-.34***	-.19***	-.20***
Study 1b	-.36***	-.35	-.34***	-.36***	-.17***	-.15*
Overconfidence						
Study 1a	.32***	.31***	.31***	.31***	.20***	.22***
Study 1b	.33***	.32***	.33***	.33***	.19***	.19***
Discrimination[a]						
Study 1a	-.16**	-.16**	-.16**	-.16**	-.12^	-.16**
Study 1b	-,16**	-.16**	-.16**	-.16**	-.14*	-.12*

All tests two-tailed: ^ $p < .10$ *$p < .05$ **$p < .01$ ***$p < .005$

Table 3. Association between automatic (A) and controlled (C) judgmental components and outcome measures (Study 1).

	Task			
	Belief Bias		Denominator Neglect	
Measure	A	C	A	C
System 1 Biases				
CRT				
Study 1a	.06	.52***	-.00	.39***
Study 1b	-.07	.30***	-.09	.42***
Belief Bias				
Study 1a			-.04	-.38***
Study 1b			-.06	-.43***
Denominator Neglect				
Study 1a	-.04	.33***		
Study 1b	-.00	.33***		
Oddball Beliefs				
Medical Myths				
Study 1a	-.04	-.23***	.04	-.19***
Study 1b	.06	-.24***	.16**	-.28***
Conspiracy Theories				
Study 1a	-.01	-.25***	.01	-.20***
Study 1b	.06	-.22***	.13*	-.23***
Paranormal Beliefs				
Study 1a	-.06	-.26***	-.01	-.25***
Study 1b	.06	-.25***	.15*	-.28***
Aggregate				
Study 1a	-.04	-.29***	.02	-.25***
Study 1b	.07	-.25***	.16**	-.29***
Knowledge Corruption				
Confidence				
Study 1a	.03	.16*	.01	.15*
Study 1b	.08	.11	.03	.06
Accuracy				
Study 1a	.04	.40***	-.05	.37***
Study 1b	-.03	.24***	-.11^	.34***
Overconfidence				
Study 1a	-.02	-.32***	.06	-.29***
Study 1b	.10	-.16**	.08	-.31***
Discrimination[a]				
Study 1a	-.04	.16*	.05	.06
Study 1b	-.11^	.07	-.05	.06

All tests two-tailed: ^ $p < .10$ *$p < .05$ **$p < .01$ ***$p < .005$

Table 4. Correlation between controlled processing (C) and outcome measures after adjusting for JTC plus JTC/Outcome indirect effects passing through C (Studies 1a and 1b).

	C based on			
	Belief Bias Task		Denominator Neglect Task	
Measure	r_{co}	Indirect Effect	r_{co}	Indirect Effect
System 1 Biases				
CRT				
Study 1a	-.50***	.65 [.42, .85]	-.37***	.30 [.15, .47]
Study 1b	-.34***	.29 [.18, .42]	-.39***	.35 [.21, .51]
Belief Bias				
Study 1a			-.31***	.33 [.14, .54]
Study 1b			-.43***	.65 [.43, .92]
Denominator Neglect				
Study 1a	-.34***	.31 [.18, .46]		
Study 1b	-.36***	.34 [.21, .48]		
Oddball Beliefs				
Medical Myths				
Study 1a	-.19***	.10 [.04, .18]	-.18**	.06 [.01, .13]
Study 1b	-.31***	.17 [.10, .25]	-.30***	.17 [.09, .26]
Conspiracy Theories				
Study 1a	-.21**	.08 [.03, .15]	-.17**	.06 [.01, 12]
Study 1b	-.26***	.13 [.07, .19]	-.23***	.12 [.06, .20]
Paranormal Beliefs				
Study 1a	-.23***	.10 [.04, .16]	-.24***	.06 [.02, .13]
Study 1b	-.28***	.13 [.08, .19]	-.28***	.14 [.08, .21]
Aggregate				
Study 1a	-.25***	.09 [.05, .15]	-.23***	.06 [.02, .13]
Study 1b	-.31***	.14 [.09, .21]	-.30***	.14 [.09, .15]
Knowledge Corruption				
Confidence				
Study 1a	.15*	-.20 [-.42, -.04]	.15*	-.15 [-.32, -.01]
Study 1b	.08	*-.41 [-1.17, .34]*	.06	*-.41 [-1.17, .34]*
Accuracy				
Study 1a	.35***	-.61 [-.93, -.33]	.36***	-.45 [-.70, -.20]
Study 1b	.27***	-2.12 [-3.10, -1.25]	.33***	-2.72 [-4.19, -1.53]
Overconfidence				
Study 1a	-.27***	2.04 [0.93, 3.42]	-.28***	1.51 [0.62, 2.77]
Study 1b	-.21***	1.65 [0.82, 2.58]	-.28***	2.30 [1.25, 3.69]
Discrimination[a]				
Study 1a	.16*	-1.07 [-2.13, -0.24]	.11	*-.12 [-.87, .63]*
Study 1b	.14*	-1.11 [-2.11, -0.21]	.05	*-.50 [-5.31, 1.14]*

Note: Indirect effects in regular font are significant; those in italics are not.

All tests two-tailed: ^ $p < .10$ *$p < .05$ **$p < .01$ ***$p < .005$

Table 5. Zero-order correlations between measures taken in Study 2.

Variable	Mean	*SD*	Zero-order Correlations 1	2	3	4	5	6	7
1. JTC[a]	0.91	0.63							
2. Accuracy (% correct)	.58	.10	-.25*						
3. Confidence (%)	63.2	16.1	.23*	.08					
4. Overconfidence	5.5	18.2	.34***	-.47***	.84***				
5. Need for Cognition	19.86	26.3	.19^	.13	-.04	-.11			
6. Need for Closure	55.9	20.0	.03	.004	.06	.05	-.18^		
7. Ammons	42.9	4.4	-.27**	.14	-.24*	-.29***	.14	-.20^	

All tests two-tailed: ^ $p < .10$ *$p < .05$ **$p < .01$ ***$p < .005$

Table 6. Trends Over Time, JTC, and the intervention for confidence in Studies 2 – 4.

Variable	Measure			
	b	se_b	p	η_p^2
		Study 2		
JTC	5.84	2.54	.024	.05
Time Trends				
Linear	-.02	.05	*ns*	<.005
Quadratic	-.01	.001	< .001	.41
Cubic	.0004	.0001	< .001	.25
Linear*JTC	-.10	.09	*ns*	.01
Quadratic*JTC	.005	.002	.038	.04
Cubic*JTC	-.00002	.0001	*ns*	<.005
		Study 3		
JTC	5.13	2.22	0.023	.06
Time Trends				
Linear	.02	.05	*ns*	<.005
Quadratic	-.01	.001	<.001	.31
Cubic	.0002	.0001	<.001	.13
Linear*JTC	-.14	.07	0.053	.04
Quadratic*JTC	.004	0.002	0.041	.05
Cubic*JTC	-.0001	.0001	*ns*	.01
		Study 4		
Time Trends				
Linear	.10	.04	.024	.04
Quadratic	-.01	.001	<.001	.21
Cubic	.0002	.0001	.002	.08
JTC	2.88	1.89	*ns*	.02
Intervention	-5.18	2.35	.03	.04
JTC*Intervention	.66	3.49	*ns*	<.005
Linear*JTC	-.08	.07	*ns*	.01
Quadratic*JTC	.002	.002	*ns*	.01
Cubic*JTC	-.00003	.00009	*ns*	<.005

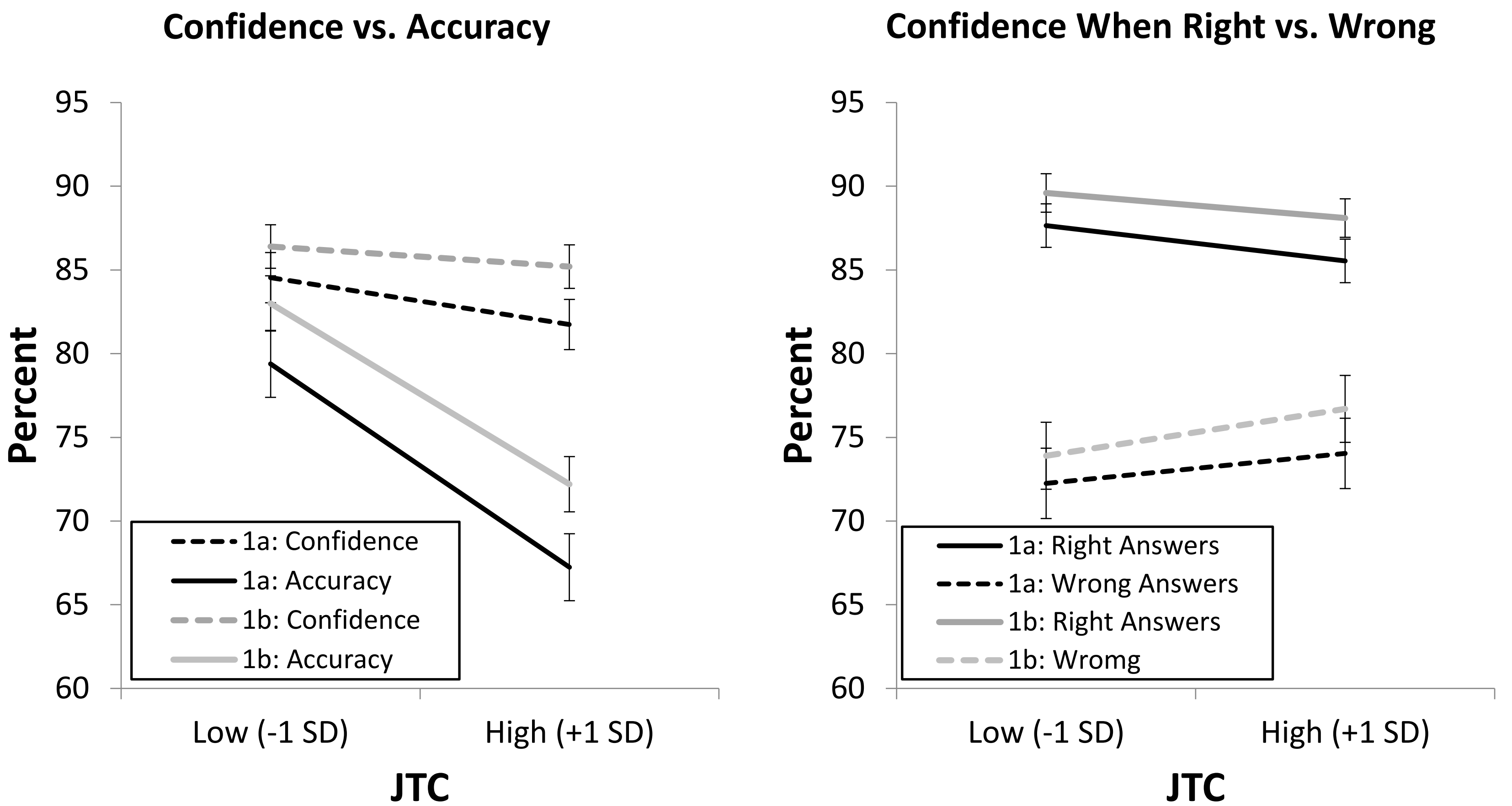


Figure 1. Confidence and accuracy on the civics quiz as a function of high and low JTC in Studies 1a and 1b (left panel). Confidence associated with right and wrong answers on the quiz as a function of high and low JTC in Studies 1a and 1b (right panel).

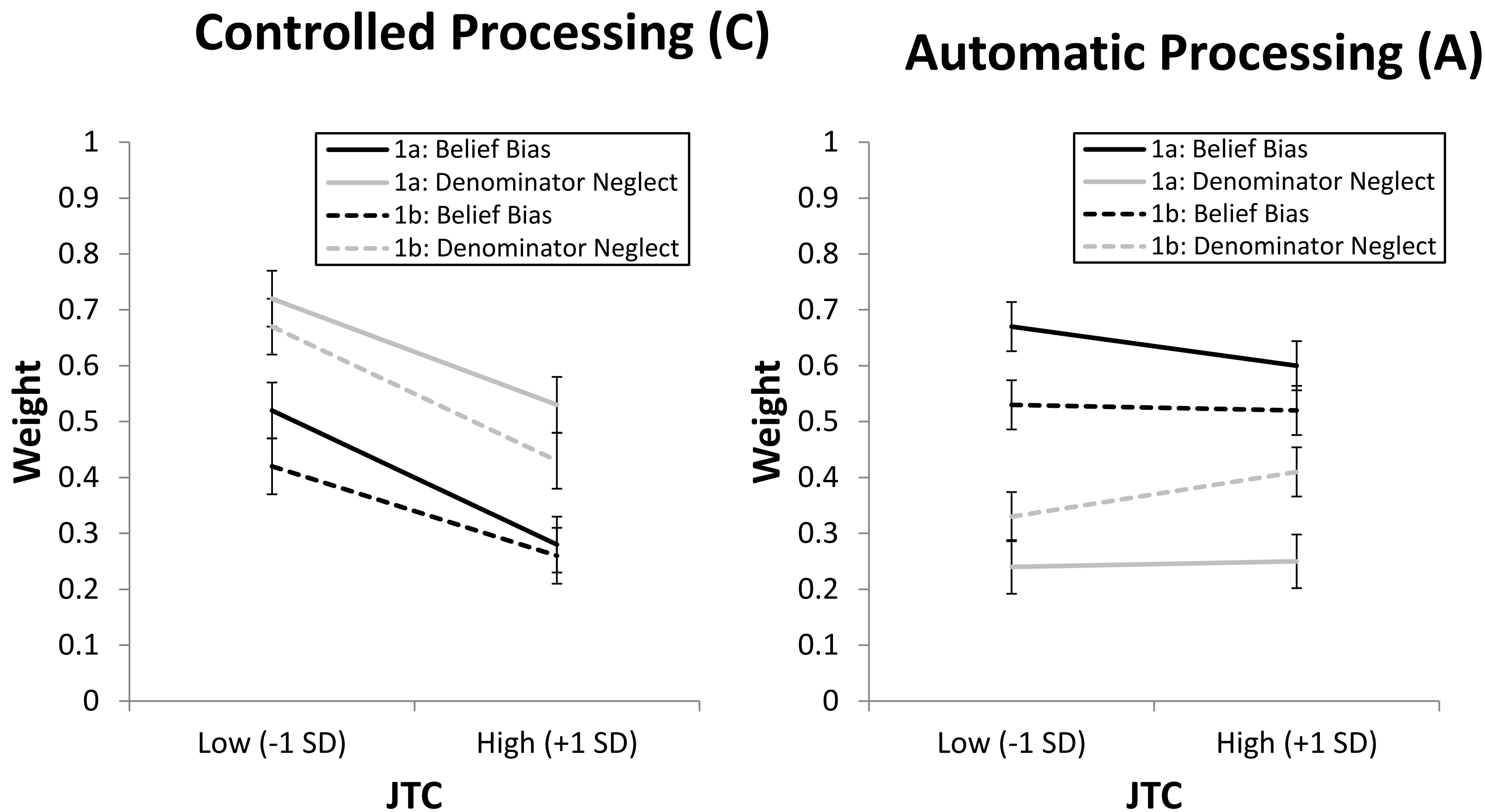


Figure 2. Weight given to automatic (*A*) and controlled (*C*) processes as a function of JTC behavior and task (Belief Bias vs. Denominator Neglect). Left panel depicts C processing. Right panel depicts *A* processing. 1a refers to Study 1a; 1b to Study 1b

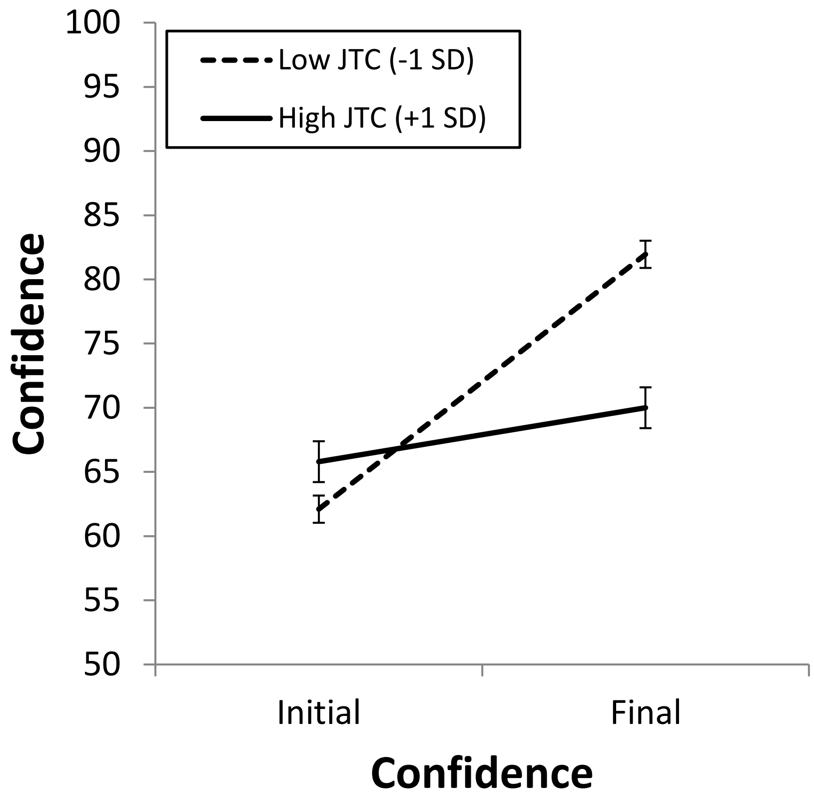


Figure 3. Average initial and final confidence expressed in the fishing task as a function of JTC in Study 1b.

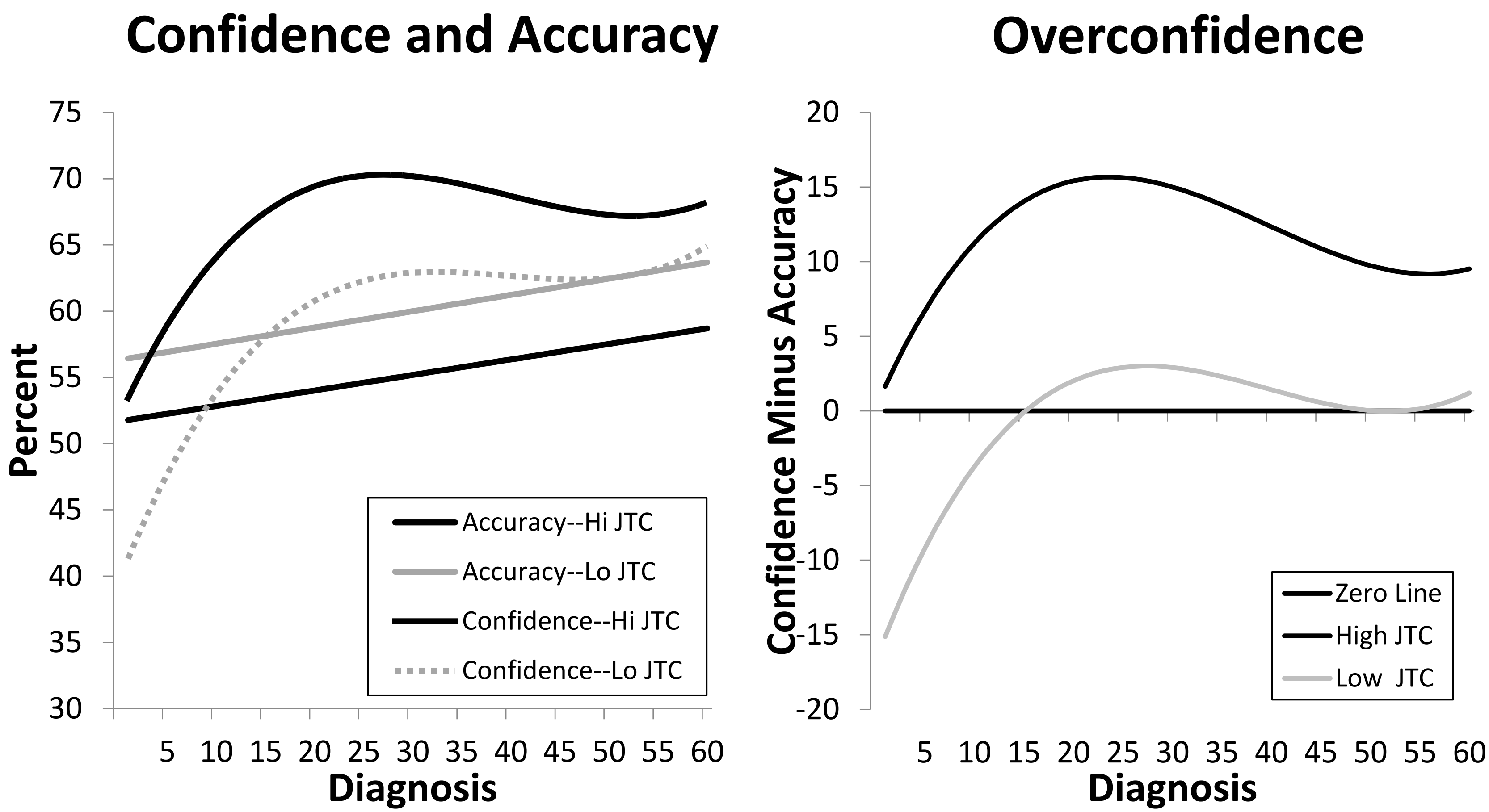


Figure 4. Time Course (Study 2) of Confidence, Accuracy, and Overconfidence for High (+1SD) and Low (-1SD) JTC participants. In left panel, dotted lines indicate confidence, solid lines represent accuracy. In right panel, the depiction comprises fitted values for confidence minus fitted values for accuracy. In both panels, grey lines represent low JTC participants, black lines represent high JTC participants.

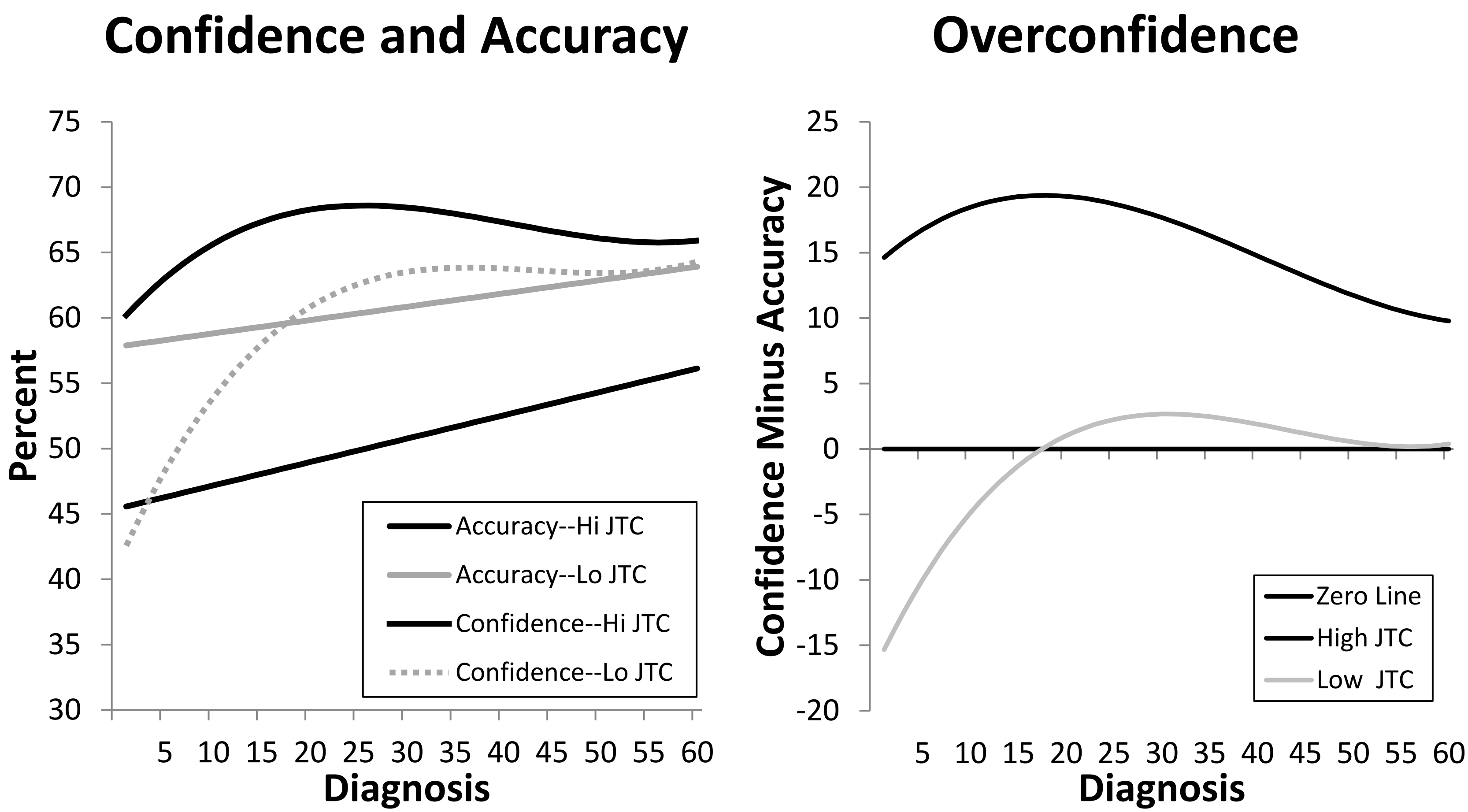


Figure 5. Time Course (Study 3) of Confidence, Accuracy, and Overconfidence for High (+1SD) and Low (-1SD) JTC participants. In left panel, dotted lines indicate confidence, solid lines represent accuracy. In right panel, the depiction comprises fitted values for confidence minus fitted values for accuracy. In both panels, grey lines represent low JTC participants, black lines represent high JTC participants.

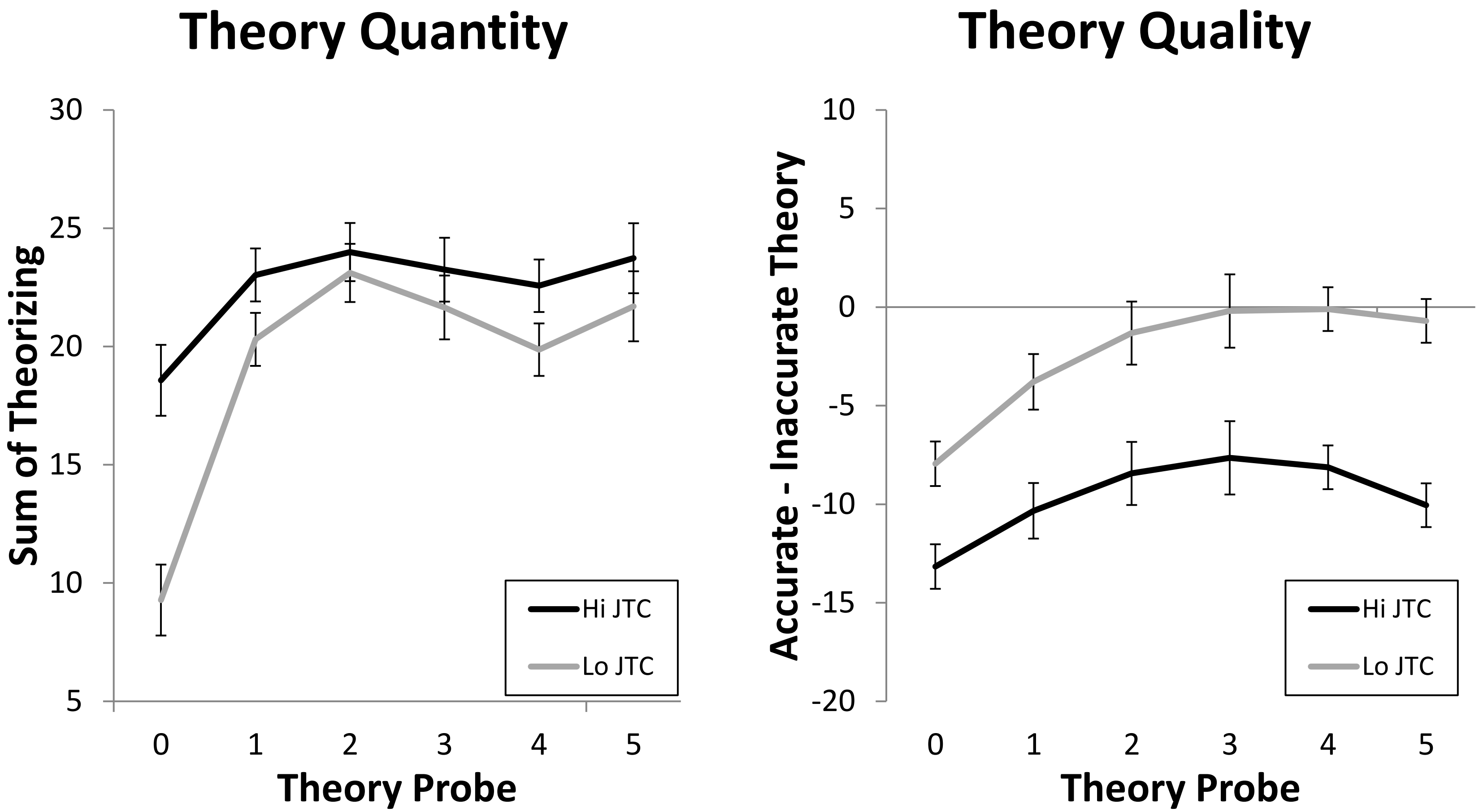


Figure 6. Time Course (Study 3) of Theorizing for High (+1SD) and Low (-1SD) JTC participants. Panels depict model fits from regression analysis. In both panels, grey lines represent low JTC participants, black lines represent high JTC participants.

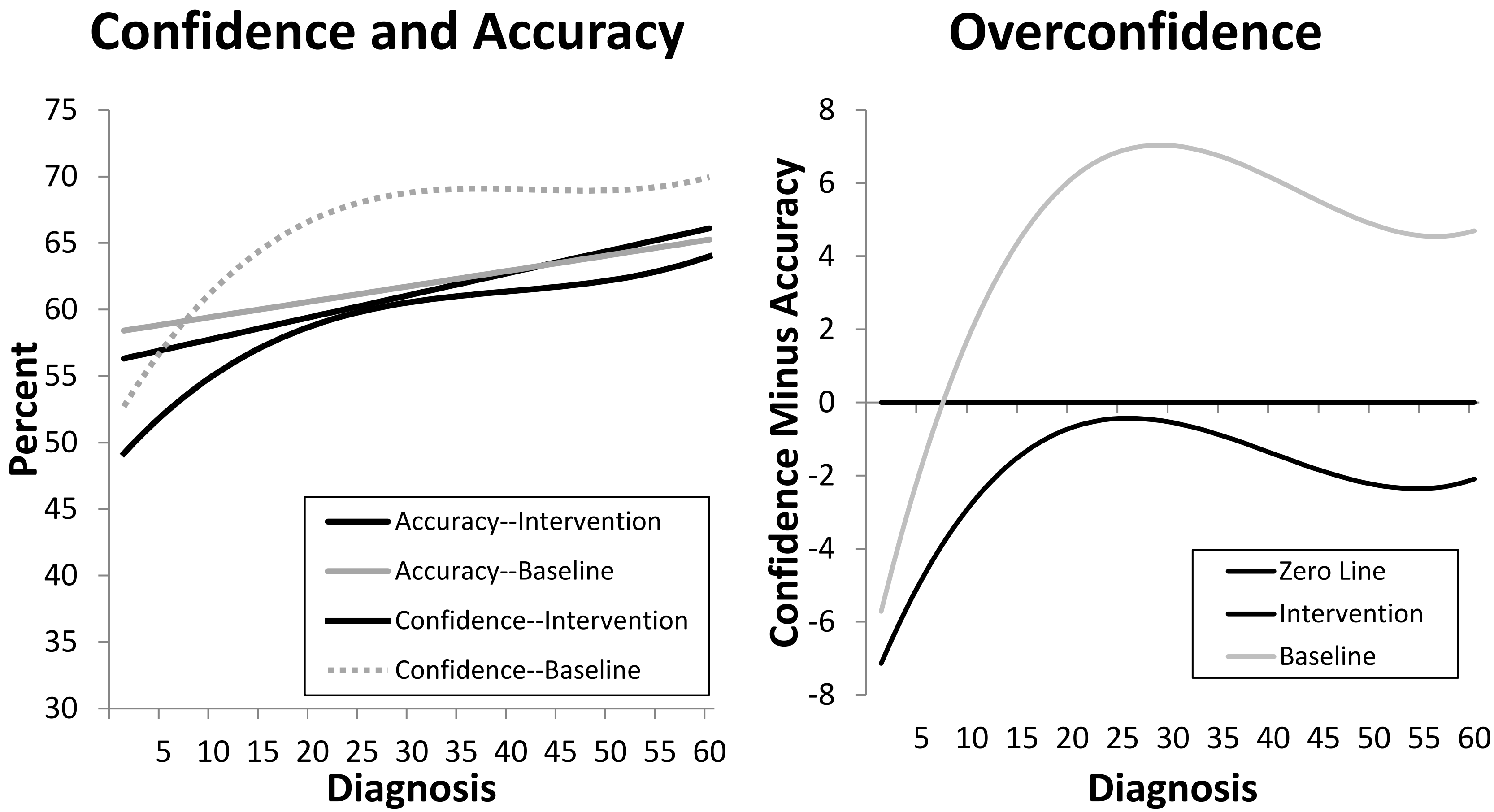


Figure 7. Time Course (Study 4) of Confidence, Accuracy, and Overconfidence intervention versus baseline participants. In left panel, dotted lines indicate confidence, solid lines represent accuracy. In right panel, the depiction comprises fitted values for confidence minus fitted values for accuracy. In both panels, grey lines represent baseline participants, black lines represent intervention participants.

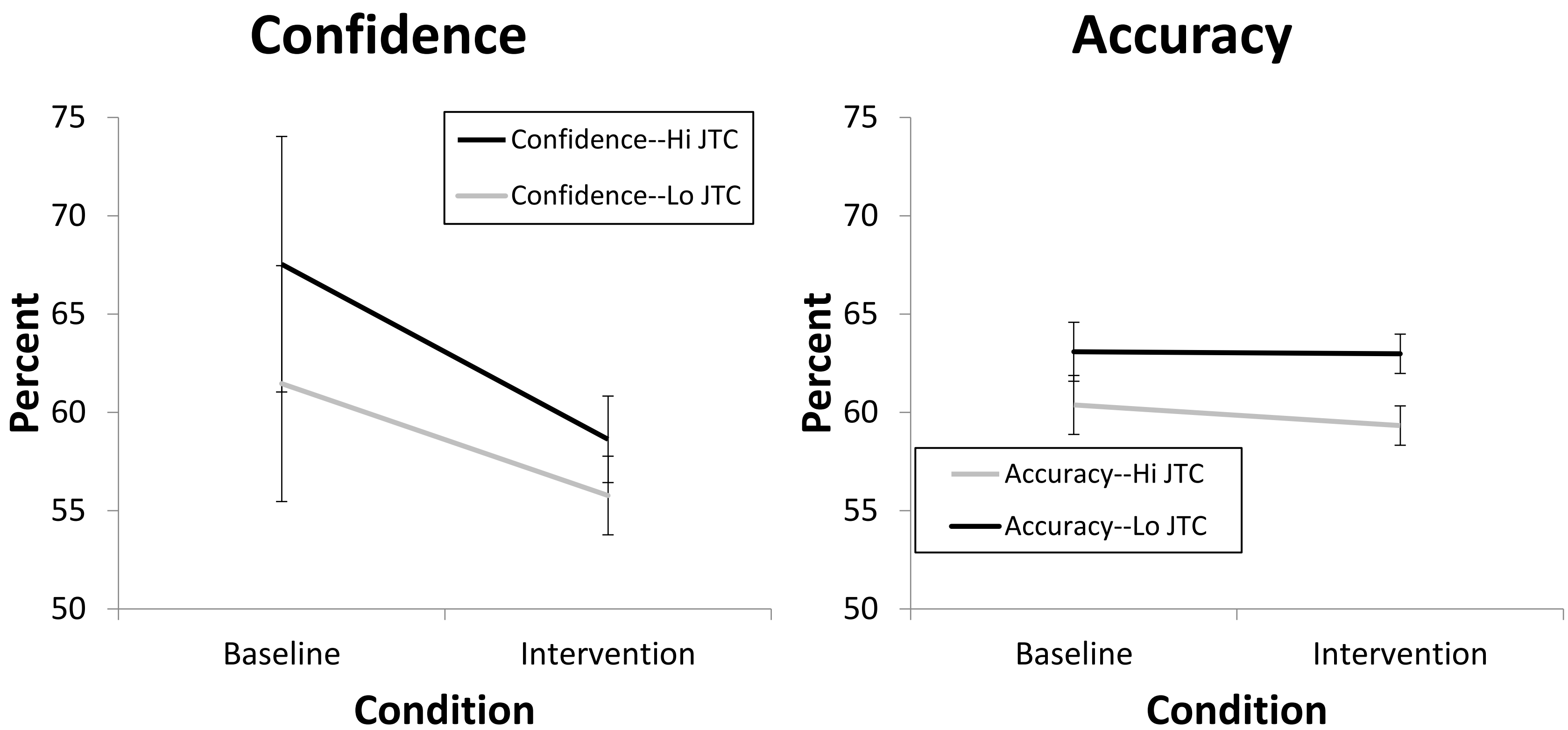


Figure 8. Effect of Intervention (Study 4) on Confidence and Accuracy for High (+1SD) and Low (-1SD) JTC participants. Panels depict model fits from regression analysis. Left panel depicts confidence; right panel depicts accuracy. In both panels, grey lines represent low JTC participants, black lines represent high JTC participants.

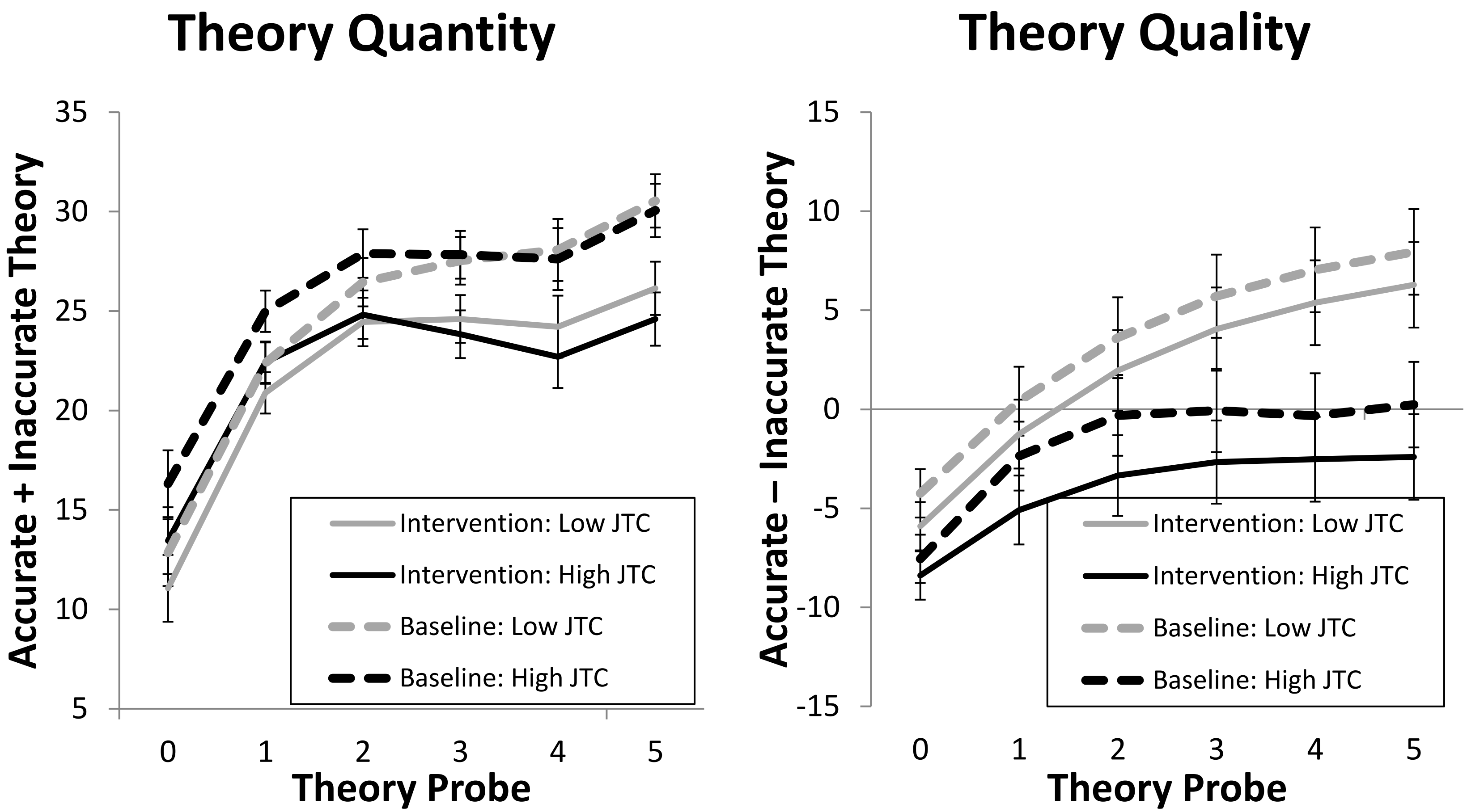


Figure 9. Time course of theorizing for participants in intervention and baseline groups (Study 4). Panels depict model fits from regression analysis. In both panels, grey lines represent baseline participants, black lines represent intervention group participants.

Chapter 5: General Discussion

Who is most likely to hold unrealistic self-views? These results suggest that it isn't rank beginners with no knowledge. Instead, it is those with a little learning.

Across three streams of research, I found that the time course of decisions plays a role in self-perceptions of ability as people are learning. In one line of research, I have mapped out the trajectory of performance, confidence, and overconfidence as people engage in novel probabilistic tasks. I have found that people generally start with rather modest self-assessments, with their confidence tracking performance rather well, but then a problem develops. With just a little learning, confidence sky-rockets far above accuracy, a phase I refer to as the "beginner's bubble" of overconfidence.

There are several reasons for the development of this beginner's bubble. First, overconfidence develops because people formulate faulty theories about how to approach a task, and once people have an idea about how to perform a task, even if it is wrong, it produces overconfidence (Chapter 2). Second, when people first engage in a task, they pay close attention to feedback to guide their decisions and are appropriately calibrated about their abilities. However, at a certain point, the experienced learner no longer carefully pays attention to feedback to discover why they made mistakes. In fact, overconfidence is not driven by feedback. It is best explained by how much time people spend making decisions. The mere act of making a series of decisions, when similar ones have been made before, drives overconfidence (Chapter 2).

In exploring mechanisms that lead to the development of the beginner's bubble, we have also identified individual differences in overconfidence resulting from a specific behavior, jumping-to-conclusions. Those who jump-to-conclusions are defined as those who collect less information to reach conclusions in problem solving tasks.

People that jump-to-conclusions have divergent trajectories of overconfidence as they learn. They have an elevated confidence curve as they are learning, lower levels of overall performance, and are thus more overconfident. These individuals are also more likely to endorse false beliefs, like conspiracy theories (Chapter 4). In this line of work, I have also modified metacognitive interventions that are ordinarily designed to reduce delusional beliefs, to quell overconfidence in a probabilistic learning task, without negatively impacting learning.

It is important to note that this work has several limitations. In the experiments, participants received perfect feedback after each trial. In life, consistent feedback like this is often unavailable. Also, in the tasks I used I traced how confidence changed as people learned truly novel tasks. There are plenty of tasks people learn in which they can apply previous knowledge to the new task. I do not know how confidence would change in these situations. Relatedly, I cannot be certain what would happen to overconfidence after the 60th trial.

Areas for Future Research

In terms of the overall pattern of overconfidence, there are two peaks in confidence—one with a little learning and another at the end of the task. There may be different psychological mechanisms that uniquely explain one peak versus the other that have yet to be explored.

Furthermore, others have found that there are different trajectories of learning and these different patterns in learning may result in different patterns in confidence or overconfidence. Not all probabilistic tasks necessarily follow a linear function of learning (Gottlieb et al., 2013). Perhaps, in many tasks learning does not need to be present for confidence to develop, and the development of confidence and overconfidence occurs from just merely being exposed to the task.

Other task variants. In many of the tasks I used throughout these studies, there is a clearly defined rule that participants must learn in a medical diagnosis task. However, there are times when diseases mutate. When the rules change in the middle of the task, diagnosis is tricky. It could be that in these types of tasks there is a different trajectory of learning and confidence.

Delay in the application of knowledge. In Chapter 2, I show how perceptions of financial literacy follow a pattern of the beginner's bubble across the lifespan. Specifically, there are some who, for whatever reason, learn harder financial lessons than others, namely those that ultimately file personal bankruptcy. These people have so much financial stress that the United States legal system allows them to liquidate their debts, file Chapter 7 bankruptcy, or reorganize them, and file Chapter 13 bankruptcy. Both filings require debtors to complete debtor education with the filings. The purpose of this education is to show people how to manage their credit in an appropriate manner. In essence, this serves as a lesson in financial literacy.

Unfortunately, there is a lag in the time between when a person must take debtor education classes and when the debts are discharged by the court. Which leads to another question-what happens to the trajectory of confidence when there is a delay between what is learned and what must be applied?

Others have found financial literacy classes to be generally ineffective, and when they are effective at changing financial behavior it is because the financial education occurs immediately prior to the relevant financial decision (Lusardi & Mitchell, 2014; Mandell & Klein, 2009). That is to say, this lag time may not make people more competent because when they have to apply knowledge, they may have already forgotten it. These classes may instead serve to build confidence much more than they serve to educate this vulnerable population.

To be sure, according to data from the financial capability study, those that have filed bankruptcy rate their financial knowledge as higher, and they also perform worse on a test of actual literacy (Dunning, 2014). After someone has learned something, knowledge may decay at a much faster rate than confidence. Future research should address rates of unlearning and how this relates to the trajectory of confidence.

Transfer learning. There are also many things that people need to learn for which they can use pre-existing knowledge.

In the present research, I have only looked at tasks that are truly novel. But there are a lot of things people need to learn that aren't truly novel. For example, if you buy a mac but you are used to using a PC, you can use some of your preexisting knowledge to learn the new task. Currently, I am exploring how confidence and learning change as people engage in tasks on which they can or cannot use pre-existing knowledge. I am studying this using hypothetical sales scenarios.

Given that overconfidence develops when people have a theory about a task, even if this theory is wrong (Chapter 2), those that apply pre-existing knowledge to learn a new task may approach this task with overconfidence.

Other individual differences. In addition to jumping-to-conclusions, there may be other individual differences that lead to divergent trajectories of overconfidence. For example, gender may impact how it develops. To be sure, others have noted that men tend to be more confident than women (Bengtsson, Persson, & Willenhag, 2005) and, as such, they may also be more likely to approach tasks with overconfidence.

In addition to these questions there are others that remain. In Chapter 2, does the beginner's bubble apply to different types of probabilistic learning, like social learning? In Chapter 3, I have shown that fluency relates to confidence in an experimental setting. What about a classroom setting? Do students that complete exams more quickly exhibit greater confidence in their abilities? In Chapter 4, I have related JTC to a host of maladaptive behaviors, such as belief in conspiracy theories. Perhaps there are situations in which JTC is adaptive. If so, what are these situations and circumstances? These are but a few extra questions I hope to address with future research.

Overall, my work highlights the circumstances in which people are the most likely to be overconfident, and several psychological mechanisms that produce these effects as people are learning. With that said, my studies suggest that the work of a beginner might be doubly hard. Of course, the beginner must struggle to learn — but the beginner must also guard against an illusion they have learned too quickly.

References

Bengtsson, C., Persson, M., & Willenhag, P. (2005). Gender and overconfidence. *Economics Letters*, *86*, 199-203.

Dunning, D. (2014). We Are All Confident Idiots. (2014, October, 27). Retrieved March 25, 2020, from https://psmag.com/social-justice/confident-idiots-92793

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*, 585-593.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, *52*, 5-44.

Mandell, L., & Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, *20*, 15-24.