Journal of Personality and Social Psychology
2018, Vol. 114, No. 1, 10–28
© 2017 American Psychological Association
0022-3514/18/$12.00 http://dx.doi.org/10.1037/pspa0000102

Overconfidence Among Beginners: Is a Little Learning a Dangerous Thing?

Carmen Sanchez
Cornell University

David Dunning
University of Michigan

Across 6 studies we investigated the development of overconfidence among beginners. In 4 of the studies, participants completed multicue probabilistic learning tasks (e.g., learning to diagnose "zombie diseases" from physical symptoms). Although beginners did not start out overconfident in their judgments, they rapidly surged to a "beginner's bubble" of overconfidence. This bubble was traced to exuberant and error-filled theorizing about how to approach the task formed after just a few learning experiences. Later trials challenged and refined those theories, leading to a temporary leveling off of confidence while performance incrementally improved, although confidence began to rise again after this pause. In 2 additional studies we found a real-world echo of this pattern of overconfidence across the life course. Self-ratings of financial literacy surged among young adults, then leveled off among older respondents until late adulthood, where it begins to rise again, with actual financial knowledge all the while rising more slowly, consistently, and incrementally throughout adulthood. Hence, when it comes to overconfident judgment, a little learning does appear to be a dangerous thing. Although beginners start with humble self-perceptions, with just a little experience their confidence races ahead of their actual performance.

Keywords: confidence, learning, metacognition, novices, overconfidence

Supplemental materials: http://dx.doi.org/10.1037/pspa0000102.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

> 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, Griffin, Milojkovic, & Ross, 1990; Dunning, Heath, & Suls, 2004; 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 article, 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 incom-

This article was published Online First November 2, 2017.
Carmen Sanchez, Department of Psychology, Cornell University; David Dunning, Department of Psychology, University of Michigan.

Correspondence concerning this article should be addressed to Carmen Sanchez, Department of Psychology, Cornell University, Uris Hall Ithaca, NY 14853-7601. E-mail: cjs386@cornell.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

petents," 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 article, 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 multicue 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 task that characterizes many of complex challenges people face in life (Brunswik, 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 right 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; Heider & Simmel, 1944; Risen, Gilovich, & Dunning, 2007). They can read meaningful patterns in putatively random or meaningless data (Chapman & Chapman, 1969; Guthrie, Weber, & Kimmerly, 1993; Ono, 1987; Rabin, 2002), 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 to gain experience with a 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, 2016; 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).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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, Victor, Mayhall, & 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 2-years 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 (Ericsson & Smith, 1991; Wallsten & Budescu, 1983). However, other research finds that even highly trained professionals remain overconfident (Cambridge & Shreckengost, 1978; Hazard & Peterson, 1973; Hynes & Vanmarcke, 1976; McKenzie, Liersch, & Yaniv, 2008; Moore, 1977; Neale & Bazerman, 1990; Oskamp, 1962; Von Holstein, 1972; Wagenaar & Keren, 1986). 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 insure 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 life span to see whether 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 postapocalyptic 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. Initially, lacking knowledge, participants would be quite cautious in their assessments of 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 nonlinear 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 of these 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 60 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 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 horizontal 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 prearranged 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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and participant as a random variable.[1] 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 nonlinear 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%), $t(37.0) = 3.70$, $p = .005$, $\eta_p^2 = .27$.

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 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.[2]

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 trial's overconfidence estimate as:

$$SE_{OC} = \sqrt{(SE_C^2 + SE_A^2 - 2 \times r_{AC} \times SE_C \times SE_A)}$$

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 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 27th 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 of the 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.

[1] Preliminary analyses assigning specific case profile (i.e., the 60 particular cases that participants diagnoses) 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 include case profile in our analyses.

[2] 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.

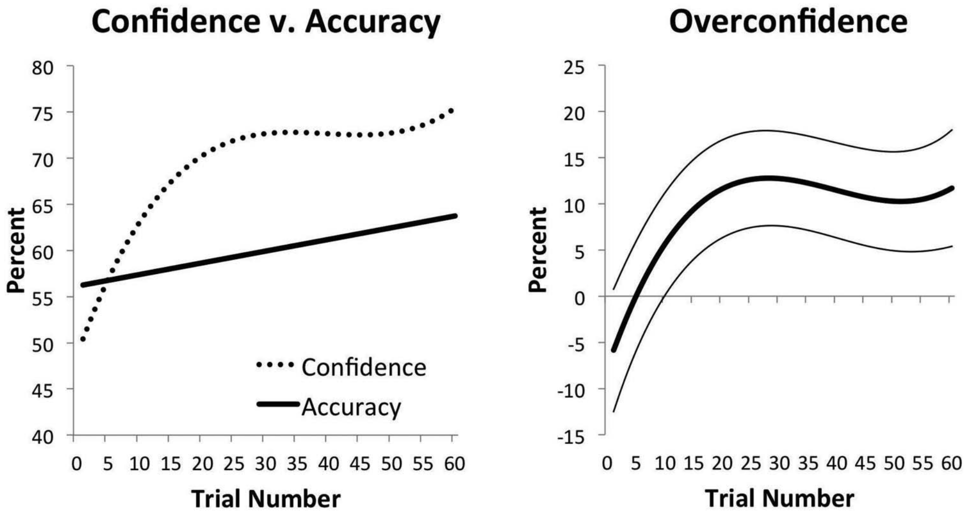

Figure 1. Confidence and accuracy trends over 60 diagnosis trials (Study 1). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

After the quiz, participants were instructed they were ready to start testing their lie detection devices. They had strapped both lie detection devices to 60 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, Thingy-mabob, 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 (Ms $= 64.2$ and 52.7, respectively, $t(49.0) = 4.62$, $p < .001$, $\eta_p^2 = .31$. 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 expressions of confidence. For confidence participants were told one of their diagnoses in the zombie task was going to be selected, and that they could win $5 depending on the accuracy of that diagnosis and the confidence they reported.

In doing so, we adopted a procedure, the Becker-DeGroot-Marschak method (Becker, DeGroot, & Marschak, 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, 2014).

This was done, in short, by telling participants they could win an additional $5 by either betting that one of their diagnoses (chosen by us) was correct or instead in a lottery. The confidence estimate they provided suggested the point at which they would shift from betting on their diagnosis to betting on the lottery. For example, if they expressed 75% confidence, that meant that they wanted to bet on their diagnosis unless the lottery odds (not yet known) just happened to provide a better than 75% chance of winning. That is, unless the lottery provided greater than a 75% chance of winning, they were saying it was more likely their diagnosis would prove accurate than they would win at the lottery. Past work has shown that this procedure prompts people to provide confidence estimates that better represent their true beliefs rather than ones contaminated by other strategies or biases, such as risk aversion or posturing (Blavatskyy, 2009; Trautmann & van de Kuilen, 2015). We constrained this "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).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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

Model	Measure				
	b	SE_b	p	η_p^2	BIC
Study 1					
Linear	.29	.045	<.001	.52	19,011
Quadratic					18,907
Linear	.29	.045	<.001	.52	
Quadratic	−.01	.002	<.001	.49	
Cubic					18,889
Linear	.06	.057	*ns*	.01	
Quadratic	−.01	.002	<.001	.49	
Cubic	.0004	.00008	<.001	.22	
Study 2					
Linear	.18	.034	<.001	.36	23,929
Quadratic					23,838
Linear	.18	.034	<.001	.36	
Quadratic	−.007	.0017	<.001	.28	
Cubic					23,810
Linear	−.05	.06	*ns*	.01	
Quadratic	−.007	.0017	<.001	.28	
Cubic	.0004	.00008	<.001	.44	
Study 3					
Linear	.30	.04	<.001	.54	24,506
Quadratic					24,355
Linear	.30	.04	<.001	.54	
Quadratic	−.01	.002	<.001	.41	
Cubic					24,289
Linear	.13	.08	*ns*	.05	
Quadratic	−.01	.002	<.001	.41	
Cubic	.0003	.0001	.004	.15	
Study 4					
Linear	.27	.046	<.001	.43	23,455
Quadratic					23,354
Linear	.27	.046	<.001	.43	
Quadratic	.006	.002	.012	.13	
Cubic					23,348
Linear	.01	.07	*ns*	.04	
Quadratic	−.006	.002	.012	.13	
Cubic	.0003	.00009	.001	.20	

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 60 patient profiles. For each, they would be 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 whether 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, participants were told that 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 at 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.

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 in any experimental session and played the additional bet, paying off those participants who won. The chance of winning the lottery was announced to be 72%. For those who expressed confidence in their diagnosis equal or greater than that, they were paid \$5 if their diagnosis was correct. For the rest, the experimenter consulted a computerized random number generator, paying the participant if the computer then generated a two-digit number (from 00 to 99) less than 72.

Results and Discussion

We performed the same analyses as we 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 a mixed model analysis, *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 substan-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

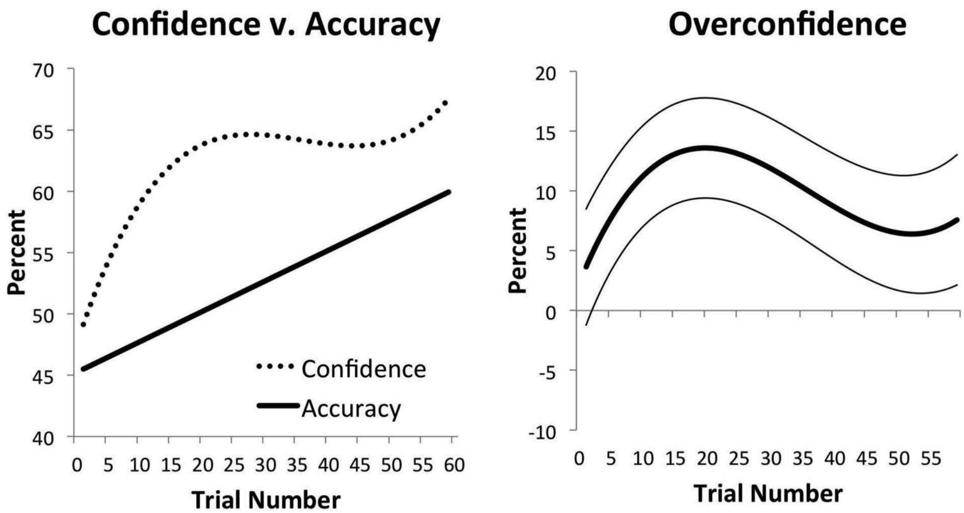

Figure 2. Confidence and accuracy trends over 60 lie detection trials (Study 2). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the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

tial 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 the previous studies and so predicted for this one.

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 12 trials. More specifically, we assessed whether participants had formed a theory about the outcome each symptom was connected to (vs. 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)

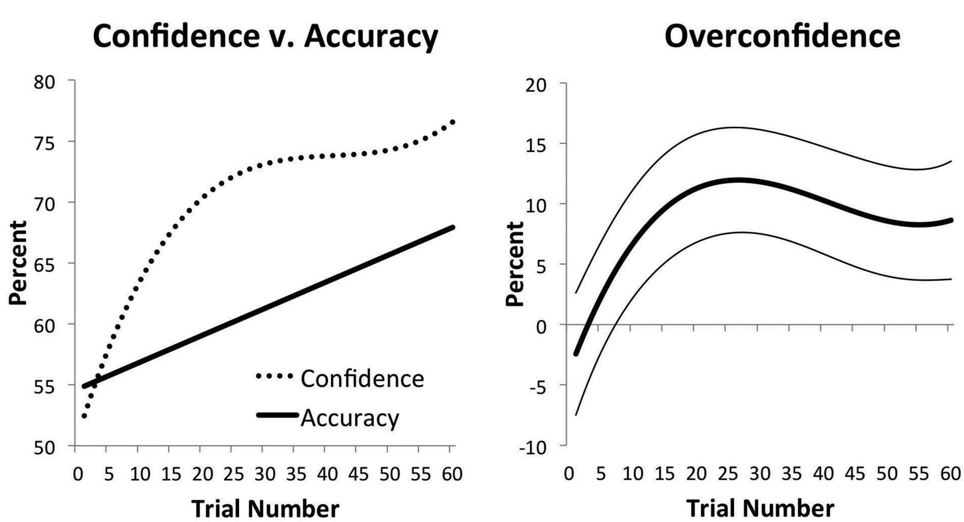

Figure 3. Confidence and accuracy trends over 60 diagnosis trials (Study 3). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

from erroneous theorizing (i.e., the participant attached the symptom to the wrong outcome). On the basis of our previous findings, 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 degree of theory development, we added a task 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 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 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. 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 devel-

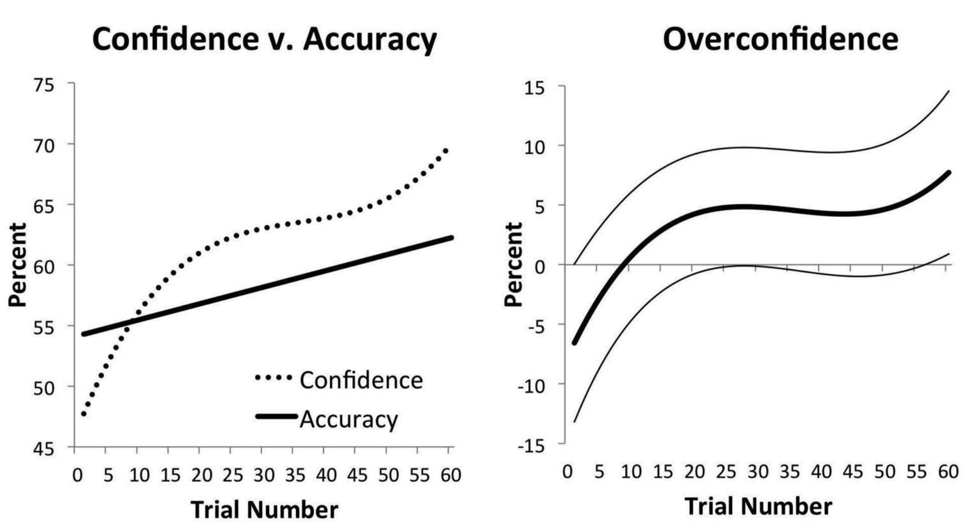

Figure 4. Confidence and accuracy trends over 60 diagnosis trials (Study 4). Left Panel: Confidence and accuracy trends. Right Panel: Overconfidence trend formed by subtracted fitted model for accuracy from fitted model for confidence. Upper and lower lines represent 95% confidence interval for the trend.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

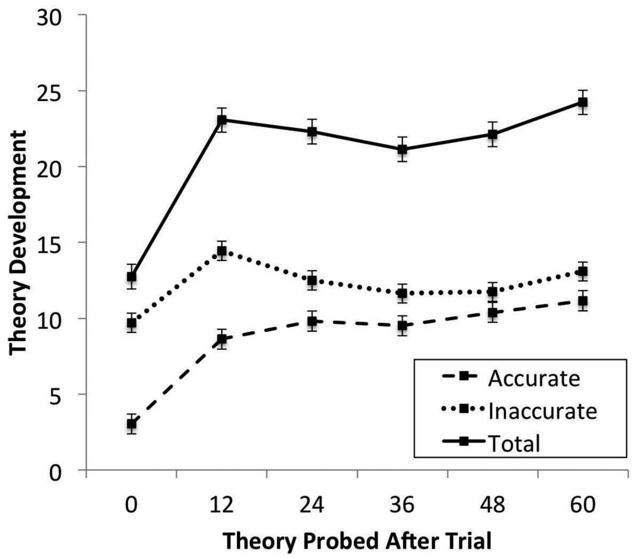

Figure 5. Theory development over experience. The figure displays theory development that is accurate, inaccurate, as well as the sum of the two.

opment 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.[3] Thus, for each theory probe, we had confidence and accuracy data that represented participant's contemporaneous performance and perception thereof. These data preserved the effects of experience we had seen previously. For accuracy, the linear component of improvement was preserved in an multilevel 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, $F(1, 46.2) = 15.41$, $p < .001$, $\eta_p^2 = .25$.[4]

We then, first, looked to see whether 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 whether 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, fulfilling the first two of the traditional steps used test for mediation, showing the independent variables (the cubic trend) predicts both the dependent measure and the mediator (Kenny, Kashy, & Bolger, 1998). The only step remaining to demonstrate 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 multilevel analysis (random slope and intercept) predicting confidence from the cubic trend adding 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 docu-

[3] One confidence judgment was omitted from subsequent analysis, lying 3.8 *SD*s away from its group mean, 1.9 *SD*s away from its nearest neighbor.

[4] 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.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

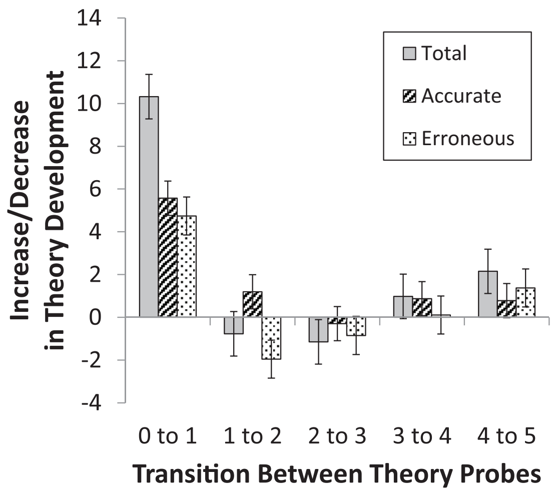

Figure 6. Degree of change in accurate, erroneous, and total theorizing taking place between theory probes.

mented the linear trend in both accurate theoretical development and diagnosis accuracy. To demonstrate mediation, we conducted a multilevel analysis on diagnostic accuracy including the linear trend in experience as well as accurate theory development as predictors. Accurate theory 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 patterning 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, Bumcrot, Lin, & Ulicny, 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 multifaceted 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 (Cunningham, 2006; Kramer, 1994; 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 life span will follow the same nonlinear 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.

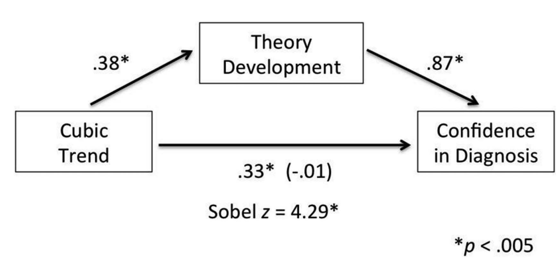

Figure 7. Mediational analysis testing whether the cubic trend in experience in theory development accounts for the cubic trend in confidence.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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 data sets, 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 7-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, postgraduate degree), and yearly income (i.e., less than $15,000, less than $25,000, less then $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 data sets to achieve a representative portrait of the United States.

Actual financial literacy. We subjected scores 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-groups variance attributable 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-groups 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-groups 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-groups 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 knowl-

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."

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

edge 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-groups 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 forward, it could only be understood backward. 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

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*	η^2	*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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

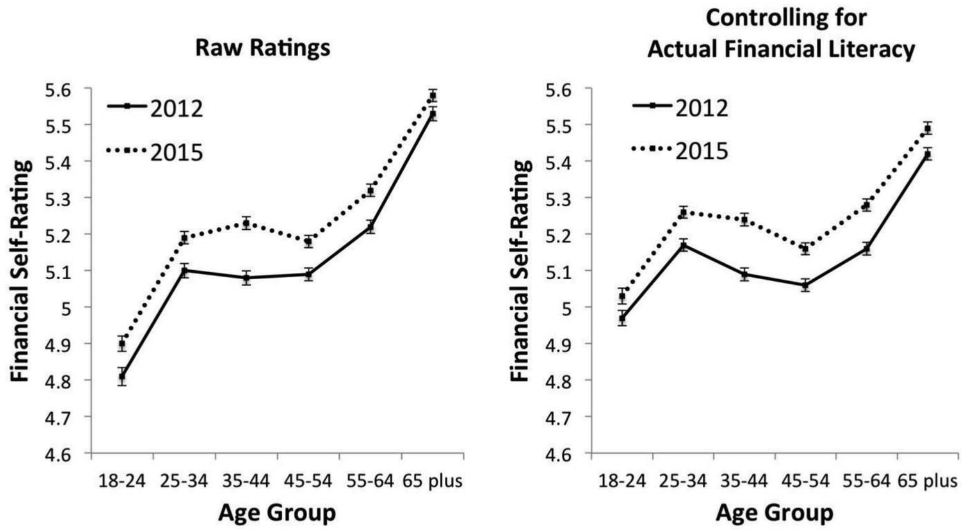

Figure 8. Self-perceived financial literacy (Studies 5a and 5b). Left Panel: Raw self-ratings. Right Panel: Self-ratings after controlling for actual financial literacy.

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 multicue 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.[5]

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.[6] 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

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.[7] 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

[5] 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 the 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.

[6] We also note that participants were so exuberant in their theorizing that 40 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).

[7] 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 that time trend, total theory development continues to predict confidence, $b = .31$, $p = .044$, $\eta_p^2 = .31$, with the time trend in confidence reduced by a marginally significant degree, Sobel $z = 1.91$, $p = .056$, two-tailed.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

found in the literature, that experts can be just as prone to overconfidence as nonexperts are, despite greater accuracy (Cambridge & Shreckengost, 1978; Hazard & Peterson, 1973; Hynes & Vanmarcke, 1976; McKenzie, Liersch, & Yaniv, 2008; Moore, 1977; Neale & Bazerman, 1990; Oskamp, 1962; Von Holstein, 1972; Wagenaar & Keren, 1986; although see contrary evidence in Ericsson & Smith, 1991; Wallsten & 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 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 life span. 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, Johnson, Banner, Dunning, & Kruger, 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 versus 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 (Hoffmann, 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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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., averse 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 preemptive 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 circumstance that might influence the relation of confidence and accuracy would be the by which why people are sorted into 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, Oudeyer, Lopes, & Baranes, 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 (Fiedler, Walther, Armbruster, Fay, & Naumann, 1996; Gill, Swann, & Silvera, 1998; Koriat, 1993, 1995; 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. http://dx.doi.org/10.1037/a0029395

Armor, D. A., & Taylor, S. E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology, 30,* 309–379. http://dx.doi.org/10.1016/S0065-2601(08)60386-X

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41,* 258–290. http://dx.doi.org/10.1037/h0055756

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. http://dx.doi.org/10.1016/j.forsciint.2010.04.016

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. http://dx.doi.org/10.1111/0022-1082.00226

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science, 9,* 226–232. http://dx.doi.org/10.1002/bs.3830090304

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. http://dx.doi.org/10.1111/jeea.12139

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine, 121*(Suppl.), S2–S23. http://dx.doi.org/10.1016/j.amjmed.2008.01.001

Billeter, D., Kalra, A., & Loewenstein, G. (2011). Underpredicting learning after initial experience with a product. *Journal of Consumer Research, 37,* 723–736. http://dx.doi.org/10.1086/655862

Blavatskyy, P. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty, 38,* 39–49. http://dx.doi.org/10.1007/s11166-008-9048-7

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review, 50,* 255–272. http://dx.doi.org/10.1037/h0060889

Cambridge, R. M., & Shreckengost, R. C. (1978). *Are you sure?: The subjective probability assessment test*. Washington, DC: 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. http://dx.doi.org/10.1257/aer.89.1.306

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. http://dx.doi.org/10.1037/h0027592

Craig, P. (2013). *The killing zone: How & why pilots die*. New York, NY: 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. http://dx.doi.org/10.1037/0033-295X.112.4.951

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, NY: Elsevier. http://dx.doi.org/10.1016/B978-0-12-385522-0.00005-6

Dunning, D. (2012). *Self-insight: Roadblocks and detours on the path to knowing thyself*. Hove, UK: 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. http://dx.doi.org/10.1037/0022-3514.58.4.568

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. http://dx.doi.org/10.1111/j.1529-1006.2004.00018.x

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. http://dx.doi.org/10.1111/1467-8721.01235

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105,* 98–121. http://dx.doi.org/10.1016/j.obhdp.2007.05.002

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416. http://dx.doi.org/10.1037/0033-295X.85.5.395

Elwin, E. (2013). Living and learning: Reproducing beliefs in selective experience. *Journal of Behavioral Decision Making, 26,* 327–337. http://dx.doi.org/10.1002/bdm.1770

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

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37–64. http://dx.doi.org/10.1037/0033-295X.83.1.37

Fetchenhauer, D., & Dunning, D. (2010). Why so cynical?: Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science, 21,* 189–193. http://dx.doi.org/10.1177/0956797609358586

Fiedler, K., Walther, E., Armbruster, T., Fay, D., & Naumann, U. (1996). Do You*Really*Know what you have seen? Intrusion errors and presuppositions effects on constructive memory. *Journal of Experimental Social Psychology, 32,* 484–511. http://dx.doi.org/10.1006/jesp.1996.0022

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. http://dx.doi.org/10.1016/j.obhdp.2005.04.003

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. http://dx.doi.org/10.1037/0096-1523.3.4.552

Fischhoff, 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.

Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G. C., Wolf, F. M., Heckerling, P. S., . . . Elstein, A. S. (2005). Do physicians know when their diagnoses are correct? Implications for decision support and error

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

reduction. *Journal of General Internal Medicine, 20,* 334–339. http://dx.doi.org/10.1111/j.1525-1497.2005.30145.x

Gill, M. J., Swann, W. B., Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75,* 1101–1114. http://dx.doi.org/10.1037/0022-3514.75.5.1101

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. http://dx.doi.org/10.1016/j.tics.2013.09.001

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435. http://dx.doi.org/10.1016/0010-0285(92)90013-R

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. http://dx.doi.org/10.1006/ceps.1993.1017

Hadar, L., Sood, S., & Fox, C. R. (2013). Subjective knowledge in consumer financial decisions. *Journal of Marketing Research, 50,* 303–316. http://dx.doi.org/10.1509/jmr.10.0518

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research, 32,* 385–410. http://dx.doi.org/10.1111/j.1468-2958.2006.00281.x

Hayward, M. L., & Hambrick, D. C. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly, 42,* 103–127. http://dx.doi.org/10.2307/2393810

Hayward, M. L., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science, 52,* 160–172. http://dx.doi.org/10.1287/mnsc.1050.0483

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. *The American Journal of Psychology, 57,* 243–259. http://dx.doi.org/10.2307/1416950

Hoffmann, 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. http://dx.doi.org/10.1037/a0037989

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, Canada: 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,* 317–340. http://dx.doi.org/10.1037/h0026818

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: 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. http://dx.doi.org/10.1111/j.1467-6494.1950.tb01260.x

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,* 401–419. http://dx.doi.org/10.1016/0022-1031(70)90052-1

Kelley, H. H., & Stahelski, A. J. (1970b). Social interaction basis of cooperators’ and competitors’ beliefs about others. *Journal of Personality and Social Psychology, 16,* 66–91. http://dx.doi.org/10.1037/h0029849

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). Boston, MA: McGraw-Hill.

Knecht, W. R. (2013). The “killing zone” revisited: Serial nonlinearities predict general aviation accident rates from pilot total flight hours. *Accident Analysis and Prevention, 60,* 50–56. http://dx.doi.org/10.1016/j.aap.2013.08.012

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. http://dx.doi.org/10.1016/j.joep.2006.11.002

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639. http://dx.doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124,* 311–333. http://dx.doi.org/10.1037/0096-3445.124.3.311

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. http://dx.doi.org/10.1037/0096-3445.126.4.349

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. http://dx.doi.org/10.1037/0022-3514.77.6.1121

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 N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2386–2388). New York, NY: 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. http://dx.doi.org/10.1002/bdm.472

McKenzie, C. R. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 771–792. http://dx.doi.org/10.1037/0278-7393.24.3.771

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,* 179–191. http://dx.doi.org/10.1016/j.obhdp.2008.02.007

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502–517. http://dx.doi.org/10.1037/0033-295X.115.2.502

Moore, P. G. (1977). The manager’s struggle with uncertainty. *Journal of the Royal Statistical Society, Series A (General), 140,* 129–165. http://dx.doi.org/10.2307/2344872

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

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

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. http://dx.doi.org/10.15369/sujms1989.2.121

Ono, K. (1987). Superstitious behavior in humans. *Journal of the Experimental Analysis of Behavior, 47,* 261–271. http://dx.doi.org/10.1901/jeab.1987.47-261

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. http://dx.doi.org/10.1037/h0093849

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261–265. http://dx.doi.org/10.1037/h0022125

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

Pope, A. (1711). *An essay on criticism*. London, England: W. Lewis.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review, 46,* 657–685. http://dx.doi.org/10.1016/S0014-2921(01)00207-0

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33,* 1492–1502. http://dx.doi.org/10.1177/0146167207305862

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., . . . Rohde, V. (2015). Unskilled unawareness and the learning curve in robotic spine surgery. *Acta Neurochirurgica, 157,* 1819–1823. http://dx.doi.org/10.1007/s00701-015-2535-0

Schotter, A., & Trevino, I. (2014). Belief elicitation in the laboratory. *Annual Review of Economics, 6,* 103–128. http://dx.doi.org/10.1146/annurev-economics-080213-040927

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, 2,* 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. http://dx.doi.org/10.1037/0278-7393.27.4.1003

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. http://dx.doi.org/10.1002/bdm.1793

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. http://dx.doi.org/10.1037/0022-3514.73.4.747

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes, 51,* 176–197. http://dx.doi.org/10.1016/0749-5978(92)90010-5

Trautmann, S. T., & van de Kuilen, G. (2015). Belief elicitation: A horse race among truth serums. *Economic Journal (London), 125,* 2116–2135. http://dx.doi.org/10.1111/ecoj.12160

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. http://dx.doi.org/10.1016/j.obhdp.2008.01.005

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110. http://dx.doi.org/10.1037/h0031322

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. http://dx.doi.org/10.1037/0022-3514.58.4.582

Von Holstein, C. A. S. S. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior & Human Performance, 8,* 139–158. http://dx.doi.org/10.1016/0030-5073(72)90041-4

Wagenaar, W. A., & Keren, G. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Maneini, & D. D. Woods (Eds.), *Intelligent decision support in process environments* (pp. 87–103). Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-50329-0_6

Wallsten, T. S., & Budescu, D. V. (1983). State of the art—Encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29,* 151–173. http://dx.doi.org/10.1287/mnsc.29.2.151

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, February 27). *The four stages of competence*. Retrieved from https://en.wikipedia.org/wiki/Four_stages_of_competence

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. http://dx.doi.org/10.1037/a0030996

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. http://dx.doi.org/10.1111/bjop.12119

Received April 1, 2017
Revision received September 8, 2017
Accepted September 9, 2017 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.