Pre-commitment and Updating Beliefs

Charles R. Ebersole

Beavercreek, Ohio

Bachelor of Arts, Miami University 2012

Master of Arts, University of Virginia, 2016

A Dissertation Presented to the Graduate Faculty of the University of Virginia

in Candidacy for the Degree of Doctor of Philosophy

Department of Psychology

University of Virginia

May, 2019

Committee

Brian A. Nosek (chair)

Timothy D. Wilson

Sophie Trawalter

Barbara A. Spellman

Abstract

Beliefs help individuals make predictions about the world. When those predictions are incorrect, it may be useful to update beliefs. However, motivated cognition and biases (notably, hindsight bias and confirmation bias) can instead lead individuals to reshape interpretations of new evidence to seem more consistent with prior beliefs. Pre-committing to a prediction or evaluation of new evidence before knowing its results may be one way to reduce the impact of these biases and facilitate belief updating. I first examined this possibility by having participants report predictions about their performance on a challenging anagrams task before or after completing the task. Relative to those who reported predictions after the task, participants who pre-committed to predictions reported predictions that were more discrepant from actual performance and updated their beliefs about their verbal ability more (Studies 1a and 1b). The effect on belief updating was strongest among participants who directly tested their predictions (Study 2) and belief updating was related to their evaluations of the validity of the task (Study 3). Furthermore, increased belief updating seemed to not be due to faulty or shifting memory of initial ratings of verbal ability (Study 4), but rather reflected an increase in the discrepancy between predictions and observed outcomes (Study 5). In a final study (Study 6), I examined pre-commitment as an intervention to reduce confirmation bias, finding that pre-committing to evaluations of new scientific studies eliminated the relation between initial beliefs and evaluations of evidence while also increasing belief updating. Together, these studies suggest that pre-commitment can reduce biases and increase belief updating in light of new evidence.

Dedications

I am indebted to a host of people who supported me during graduate school. There are more than I can name here, but I'd like to focus on a few people who I could not have done this without.

First, I'd like to thank my primary advisor, Brian Nosek. I'm a much better researcher for having worked with you. I'm thankful for our many engaging, challenging, and productive conversations about this work and other projects.

I've been very fortunate to have a fantastic group of advisors throughout my time as a researcher. I'd like to thank Tim, Sophie, and Bobbie for serving on my dissertation committee. A special thank you to Tim Wilson, who I got to watch teach several semesters. It was a privilege to work with you and learn from you. I'd also like to thank Shige Oishi for advising me throughout graduate school. Some of my favorite conversations about research happened in your cramped office with most of the social area grad students. Finally, I'd like to thank Carrie Hall who gave me my start in research at Miami University. Thank you for nurturing my curiosity in social psychology and giving me so many opportunities to pursue my interests.

I was lucky to go through grad school with many fantastic peers. In particular, I'd like to thank Jordan Axt, Calvin Lai, Nick Buttrick, and Courtney Soderberg. I learned so much from you all. I owe a special thank you to my favorite academic debate partner, Erin Westgate. I learned as much about research from our conversations as from any other source.

My family has been an incredible source of support in graduate school. Thank you to my Mom and Dad and my sister Megan for being there for me through all the highs and lows. I'm thankful to the second family I picked up during graduate school as well, my in-laws Sandy, Wayne, Chelsea, and Nick. You have all been a source of strength for me.

I'd like to thank the Fetzer Franklin Fund for supporting and funding this research. I'd also like to thank everyone who has worked on Deciphering the Decline Effect – this dissertation began as a part of that project.

Finally, the biggest thank you goes to my incredible spouse, Rachel. You've been the defining force throughout this period of my life. I can't imagine it without you. Thank you for all of the laughs, love, and wisdom.

Pre-commitment and Updating Beliefs

The ANOVA wasn't significant? You know, those two conditions look pretty similar, and that one's pretty different. Let's run a planned contrast and see if it comes out. -Charles R. Ebersole (personal communication, 2012)

Beliefs are mental representations that are thought to be true (Bogdan, 1986; Clark & Chase, 1972; Gilbert, 1991). Beliefs organize past thoughts and experiences so that they can explain current experiences and predict future events. For instance, many people believe that the Earth revolves around the sun and that the Earth's axis is slightly tilted. These features produce a weather pattern consisting of four seasons. If that person experiences extreme heat during the summer, they can explain it with their belief about the movement and orientation of the Earth: it is hot because where I live is tilted closer to the sun. They can also use this belief to make predictions about what the weather will be like in the future, giving them hope that the extreme heat will indeed end.

Updating beliefs

Many beliefs are difficult to prove or disprove (e.g., does God exist?). Many others, however, have observable evidence that have bearing on whether or not the belief is true. Someone observing the same patterns of weather year after year is likely to have great faith in their belief about the progression of seasons. Conversely, someone with misplaced faith in their baking ability may find that representation to be less true when cleaning up half-finished cookies after a party. In each case, the person in question is assessing the extent to which their belief explains their current experiences.

Multiple psychology theories attempt to describe how individuals update beliefs in light of new evidence. Information Integration Theory (e.g., Anderson, 1971) suggests that individuals

are always updating their beliefs as they encounter new information. This new evidence is weighted based on many features, such as the strength of the evidence or its source (e.g., whether the messenger has perceived expertise, Chaiken, 1980; 1987; Kelman, 1958, or in-group membership, Cohen 2003; Mackie, Worth, & Asuncion, 1990). An individual's belief is determined by the sum total of weighted evidence. Similarly, Lay Epistemology (Kruglanski, 1990; Kruglanski, Dechesne, Orehek, & Pierro, 2009) contends that individuals first form hypotheses about the world which they then seek evidence to support or confront. Once they have reached what they consider to be an acceptable amount of evidence, they form their belief. Then, if they later have reason to question the belief, they may begin the process again, seeking new information and changing their belief if necessary (Kruglanski, 1990; Kruglanski & Freund, 1983; Mayseless & Kruglanski, 1987).

While both theories argue that beliefs are subject to further testing and revision based on new evidence, several other cognitive processes make revising beliefs difficult. Cognition is influenced by goals and desires, shaping interpretations of new evidence and affecting their impact on existing beliefs (Kunda, 1990; Simon, Snow, & Read, 2004). Also, individuals tend to estimate the likelihood of an event as being higher once they know that the event has already occurred (known as hindsight bias, Campbell & Tesser, 1983; Christensen-Szalanski & Willham, 1991; Fischhoff, & Beyth, 1974; Hom & Ciaramitaro, 2001). This bias has multiple sources (Nestler, Blank, & Egloff, 2010; Roese & Vohs, 2012). Once an event has occurred, its occurrence seems more inevitable and constructing a narrative for how the event came to be becomes metacognitively easier (Hawkins & Hastie, 1990; Sanna & Schwarz, 2006). Hindsight bias can also protect self-esteem, making events seem more predictable or more in line with one's beliefs or desires (Campbell & Tesser, 1983). Finally, when seeking evidence that has

bearing on beliefs, individuals are likely to search for information that confirms their beliefs as opposed to challenges them (known as confirmation bias, Nickerson, 1998; Wason, 1966; 1968), unless they are dissatisfied with the predictions produced by their current beliefs and wish to disconfirm them (e.g., a negative medical prognosis, Ditto & Lopez, 1992). Individuals also weigh new evidence in light of their beliefs and experiences. For instance, lay people and scientists alike judge research that supports their beliefs as more valid and scientifically sound than research that conflicts with their beliefs (Bastardi, Uhlmann, & Ross, 2011; Hergovich, Schott, & Burger, 2010; Lord, Lepper, & Ross, 1979; Mahoney, 1977). These motivational and cognitive biases reduce the likelihood that individuals will update their beliefs in light of disconfirming evidence – such evidence may be interpreted in ways that privilege the current belief or may simply never be sought out.

A problem of timing? Interpreting evidence while updating beliefs

Motivated cognition and biases like hindsight and confirmation bias lead people to discount evidence that conflicts with their desires and prior beliefs (Bastardi, Uhlmann, & Ross, 2011; Kunda, 1990; Lord, Lepper, & Ross, 1979; Mahoney, 1977; Simon, Snow, & Read, 2004). However, to know that evidence conflicts with their beliefs, people must know the results of that evidence. This requires individuals to interpret evidence (Is this what I would have predicted? Is this a good test of my belief?) at the same time that they are updating their beliefs (Does this challenge my beliefs? Should I change my belief?). If results conflict with beliefs, individuals may be motivated to alter their interpretations of the new evidence. These biased interpretations would likely challenge existing beliefs less and reduce the amount that people update their beliefs.

As an example, consider an aspiring baker. This person believes in his baking ability and is considering devoting more time and money to his hobby. However, before buying that expensive new stand mixer, he wants to know that people really do enjoy his baking. He decides to bring a batch of his newest cupcakes to an office party and observe how many of them are eaten. At the end of the night, half of the cupcakes are gone.

How might the baker interpret this new evidence? There could be several explanations for why only half of his cupcakes were eaten. Maybe his office mates tend to prefer cookies to cupcakes. Maybe these cupcakes were not the best representation of his baking skills; maybe cakes are his specialty. Maybe his cupcakes were so rich and delicious, his coworkers did not need a second one. With all of this in mind, maybe a half-eaten plate of cupcakes is what the baker should have expected all along.

Through this thought process, the baker has constructed a postdiction (a prediction formed after results are known) that he can use to interpret the results of his study. That postdiction may match what he would have predicted before knowing the outcome of his test. However, if the outcome is worse than he would have actually predicted, his postdiction may bias toward the results, likely reconciling those results with his existing beliefs (Campbell & Tesser, 1983; Christensen-Szalanski, & Willham, 1991; Hom & Ciaramitaro, 2001). Furthermore, this explanation of events may be easier to construct given his knowledge of the results (Hawkins & Hastie, 1990; Sanna & Schwarz, 2006). Finally, he may not even realize the discrepancy between his postdiction and what he would have predicted (Pronin, Gilovich, & Ross, 2004; Pronin, & Kugler, 2007; Pronin, Lin, & Ross, 2002). This may lead him to interpret the results as being fairly consistent with his beliefs.

Separating predictions and judgments of evidence from learning results

If knowledge of results leads people to alter their predictions and judgments of new evidence, then pre-committing to predictions and judgments, before knowing new results, may reduce biased interpretations. Without knowledge of the outcomes, predictions cannot be biased by them. The act of pre-commitment, then, may be an opportunity to clearly separate what was actually believed prior to knowing the outcome from the biased reconstruction of prior beliefs based on actually knowing the outcome. This could decrease opportunity to employ reasoning biases, or make them more obvious and likely to be discounted because the pre-commitments are known and unavoidable. To the extent that postdictions bias toward observed results, pre-commitment should result in disconfirming outcomes that might be more impactful on beliefs.

From the perspectives of Information Integration Theory (Anderson, 1971) and Lay Epistemology (Kruglanski, 1990; Kruglanski, Dechesne, Orehek, & Pierro, 2009), more challenges to beliefs may lead to more belief updating. By preventing post-hoc adjustments of predictions, pre-commitment will likely produce evidence that is more inconsistent with current beliefs. If this evidence is being incorporated into current beliefs, individuals should update their beliefs more than they would if allowed to construct postdictions. In both cases, individuals would be responding to the discrepancy between their prediction and the observed results; pre-commitment just increases this discrepancy by preserving unbiased predictions.

A motivated-reasoning escape hatch?

However, from the perspectives of motivated reasoning and cognitive consistency theories, these challenges may produce more resistance to evidence and inhibit belief updating. Individuals are motivated to resolve inconsistencies among their beliefs, especially those that reflect on their self-image (Festinger, 1957; Gawronski, 2002; Greenwald, 1980; Greenwald & Ronis, 1978; Heider, 1958). Hindsight bias provides a simple way of resolving, or at least

reducing, inconsistencies between beliefs and new evidence (Campbell & Tesser, 1983; Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988). This may lessen the threat that new evidence poses to beliefs and self-esteem. Without this route, however, individuals must contend with inconsistencies in other ways. For instance, they may seek to undermine the validity of the new evidence, more closely examining how it was produced and looking for possible flaws or weaknesses (e.g., Ditto, Munro, Apanovitch, Scepansky, & Lockhart, 2003; Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998). Whereas individuals who do not pre-commit to predictions might accept an artificially reduced gap between their postdiction and results, individuals who do pre-commit may face a greater threat to their beliefs and self-concepts. This may motivate them to find another cognitive escape hatch for rejecting or reducing the impact of new evidence.

Overview of studies

I examined these competing possibilities in a series of studies. In an initial pair of studies (Studies 1a and 1b), I examined whether participants updated their beliefs about their verbal abilities more or less after pre-committing to predictions about performance on a difficult anagrams task. I also examined whether or not participants needed direct experience with disconfirming evidence (Study 2) and/or needed the task to be described as diagnostic of verbal ability (Study 3) in order for pre-commitment to have an effect on belief updating. I then investigated whether participants could accurately recall their initial beliefs or if their memory of their initial beliefs had shifted toward their final beliefs (Study 4). Finally, I examined a mechanism for the effect of pre-commitment on belief updating (Study 5) and tested pre-commitment as an intervention to reduce biased evaluations of scientific evidence (Study 6).

Studies 1a and 1b - Testing the effect of pre-commitment on belief updating

Does pre-commitment facilitate belief updating or motivate discounting of new evidence? In two initial studies, participants rated their verbal ability before and after completing a difficult anagrams task. Participants either pre-committed to a prediction of how long they would need to complete the task or reported their postdictions about performance. Finally, participants answered several questions about the diagnosticity of the task for measuring verbal ability. If pre-commitment facilitates belief updating, those who pre-committed to predictions should update their beliefs more than those who report postdictions. However, if pre-commitment encourages motivated reasoning, participants may update their beliefs in similar amounts, while also denigrating the task as a cognitive escape hatch for confronting their beliefs.

Methods

Participants

Participants were recruited from two large online panels. From each, I planned to collect 1,500 participants who passed a series of attention check questions.[1] This sample size was determined by constraints in access to the panel, but would provide adequate power to detect even small effects (80% power to detect d = .14). In total, because of the sampling pace by the panel provider, 1,644 participants were collected from the first panel and 1,550 were collected from the second panel that met the inclusion criteria. The samples were predominantly middle aged (M = 52.37, SD = 13.52; M = 43.20, SD = 17.19), female (72.4%; 58.7%), and White (86.1%, 70.8%). The second sample was a planned direct replication of the first sample; participants in both samples completed the same procedure. The results of both studies were highly similar, so they are combined, below, for simplicity.

[1] Exclusion criteria can be found at https://osf.io/a8n2b/.

Materials and Procedure

Participants were first introduced to the construct of verbal ability. They read: *Verbal ability is an important quality. Verbal ability includes essentials like spelling, grammar, and reading comprehension. These core skills can combine to predict people's ability on many tasks related to success in school, the workplace, and life in general.* Participants then rated their verbal ability, relative to other people, on an 11-point scale (1 - *terrible, among the very worst*, 2 - *very much worse than average*, 3 - *much worse than average*, 4 - *somewhat worse than average*, 5 - *slightly worse than average*, 6 - *equal to the average*, 7 - *slightly better than average*, 8 - *somewhat better than average*, 9 - *much better than average*, 10 - *very much better than average*, 11 - *exceptional, among the very best*). Participants saw only the words associated with each scale point, not the number.

Next, participants were told that the ability to solve anagrams was diagnostic of overall verbal ability. Along with this message, participants saw a pair of anagrams, presented with their solutions (e.g., ciles, answer = slice; terny, answer = entry) and estimated how much time they would need to solve two equally difficult anagrams. Participants were randomly assigned to either record their prediction (pre-commitment condition) or simply to think about their prediction (hindsight condition). Participants in the pre-commitment condition recorded their prediction using a drop-down menu that ranged from 1 second to 120+ seconds in 5 second increments (for the purposes of analyses, scores of 120+ seconds were recoded as 125 seconds). Previous research suggests that most individuals believe they will need much less time than two minutes to solve these two anagrams. In actuality, most people will need significantly longer than two minutes (Hom & Ciaramitaro, 2001). Therefore, this created a task where participants were likely to encounter evidence that went against their predictions.

All participants were then given two minutes to solve a second pair of anagrams that were similarly difficult to the ones they saw previously (e.g., ochsa, grabe).[2] If participants did not complete the task within the allotted two minutes, the program automatically advanced them to the next page. There was a timer placed on the screen so that the participant knew how much time he or she had left to complete the task. After the anagrams task, participants in both conditions were reminded of their initial predictions – participants in the pre-commitment condition were shown the prediction that they had reported; participants in the hindsight condition were asked to remember and report their prediction (hereafter referred to as their postdiction).

All participants then reevaluated their verbal ability. Participants were told, *Direct feedback on task performance compared to expected performance can be useful for evaluating one's skills. After performing the anagram task, how would you rate your verbal ability compared to other people?* They then rerated their verbal ability on the same 11-point scale used before.

Finally, participants indicated whether or not they agreed with five statements concerning the validity of the anagrams task. The statements were: 1) "The anagrams were more difficult than the example anagrams." (response options: *yes* or *no*); 2) "The timer was distracting." (response options: *yes* or *no*); 3) "Verbal ability is not important." (response options: *yes, verbal ability is unimportant* or *no, verbal ability is an important skill*); 4) "These anagrams do not measure verbal ability." (response options: *yes, these anagrams do not measure verbal ability* or *no, the anagrams are indicators of verbal ability*); 5) "The test is too short to provide any meaningful information." (response options: *yes, the test is too short to be meaningful* or *no, the*

[2] Answers: chaos, barge

test is brief, but still meaningful).” These items were included for exploratory analyses. The number of *yes* responses were summed as a measure of discounting the validity of the task. All materials and data can be found at https://osf.io/82etx/.

Results

Confirmatory Analyses[3]

I first examined whether pre-committing to predictions led to greater updating of beliefs. As a measure of belief updating, participants’ initial verbal ability ratings were subtracted from their final verbal ability ratings (thus, negative values indicate lower ratings of verbal ability after completing the anagrams task compared to before the task). Overall, participants tended to lower their ratings of their verbal ability after completing the anagrams task, likely reflecting the fact that most people could not complete it in the allotted time (Sample 1a: M = -1.60, SD = 1.84, Sample 1b: M = -1.76, SD = 2.07). However, this trend was more pronounced among those who were required to commit to their prediction before the task, with participants in the pre-commitment condition (Sample 1a: M = -1.78, SD = 2.02, Sample 1b: M = -2.01, SD = 2.27) updating their beliefs more than those in the hindsight condition (Sample 1a: M = -1.41, SD = 1.61, Sample 1b: M = -1.54, SD = 1.85), Sample 1a: $t(1,630) = 4.10$, $p < .001$, $d = .20$, 95% CI [.11, .30], Sample 1b: $t(1,541) = 4.42$, $p < .001$, $d = .22$, 95% CI [.12, .32].

I also examined the amount of time participants predicted needing to complete the anagrams. Replicating previous research (e.g., Hom & Ciaramitaro, 2001), participants in the hindsight condition postdicted that they had thought they would need longer to complete the task (Sample 1a: M = 65.61, SD = 40.72, Sample 1b: M = 62.97, SD = 40.31) than participants in the pre-commitment condition had predicted they would need (Sample 1a: M = 39.83, SD = 31.08,

[3] Confirmatory analyses were preregistered at https://osf.io/6he7m/ (for the first sample) and https://osf.io/7dtmz/ (for the second sample).

Sample 1b: M = 40.09, SD = 33.77), Sample 1a: $t(1,635) = 14.44$, $p < .001$, $d = .74$, 95% CI [.63, .84], Sample 1b: $t(1,546) = 12.05$, $p < .001$, $d = .61$, 95% CI [.51, .72]. That is, predictions recorded *before* completing the anagrams task were shorter than postdictions reported *after* completing the anagrams.

Exploratory Analyses - Predicted vs. actual performance

The anagrams task was fairly difficult for participants. Overall, 22.1% of participants correctly answered both anagrams within the two minute time window. Very few participants (9.0%) completed both anagrams in as much or less time than they predicted they would need. Participants that solved both anagrams in under two minutes were similarly likely to be in either condition (340 in pre-commitment condition; 366 in hindsight condition), $d = .07$, 95% CI [-.07, .22], but participants were more likely to perform in line with their postdictions in the hindsight condition (N = 169) than with their predictions in the pre-commitment condition (N = 118), d = .36, 95% CI [.12, .60]. That is, participants were more likely to remember their expected performance as being consistent with their actual performance after knowing how they performed.

Finally, I examined participants' assessments of the validity of the task. Participants endorsed a similar number of statements that would invalidate the anagrams task in both conditions (pre-commitment condition: M = 1.95, SD = 1.18; hindsight condition, M = 1.93, SD = 1.15), d = -.02, 95% CI [-.09, .06]. That is, pre-commitment did not make participants more likely to discount the anagrams as being diagnostic of verbal ability after completing the task.

Discussion

Participants who reported postdictions after the task "predicted" that they would need much more time to complete the task than those who reported their predictions prior. In fact,

those who reported postdictions were more likely to be "correct" in their predictions about their performance on the anagrams task. These findings comport with the many previous demonstrations of hindsight bias. However, these studies also suggest a new benefit of pre-commitment: that it facilitates belief updating in the face of new evidence. Those who could not retroactively shift their predictions tended to update their beliefs about their verbal ability more than those who could engage in hindsight bias. Finally, participants in both conditions rated the anagrams task as similarly valid. This suggests that pre-commitment does not lead individuals to engage in other types of motivated reasoning as a way to discount new evidence.

Studies 1a and 1b suggest that individuals are incorporating new evidence (as provided by the anagrams task) into their beliefs about verbal ability (e.g., Anderson, 1971; Kruglanski, 1990). Pre-commitment, in this instance, protects predictions from bias and leads to larger gaps between predictions and results. However, it is possible that pre-commitment elicits other psychological experiences, beyond the consideration of new evidence, that would increase belief updating. For instance, pre-commitment may emphasize the fact that the participant's prediction was incorrect, irrespective of how discrepant it was from results. If this is the case, pre-commitment should still lead to increased belief updating even if participants are simply told that their predictions are likely incorrect. Conversely, if pre-commitment affects belief updating because it debiases the gap between predictions and results, participants should need to directly experience that gap in order to show the effect. I examined this possibility in Study 2.

Study 2 - Direct experience with disconfirming evidence

Methods

Participants

I initially planned to collect 1,500 participants for this study. Upon doing so and analyzing the data, I found that the primary effect from Study 1, that pre-commitment increases belief updating, did not replicate among those who completed the anagrams (the experience condition, described below). This could be evidence of the lack of reliability of that finding. However, since this effect was relatively small in Studies 1a and 1b and only half of participants in the current study completed the same procedure as those studies, it is possible that this test was underpowered. To address this, I collected an additional 1,500 participants. *P*-values are reported both uncorrected and using the *p-augmented* formula to address the contingent decision of continuing data collection after observing the initial outcomes (Sagarin, Ambler, & Lee, 2014).

The final sample was recruited from two large online panels. From each, I planned to collect 1,500 participants who passed a series of attention check questions.[4] In total, because of the sampling pace by the panel provider, 1,558 participants were collected from the first panel and 1,504 were collected from the second panel that met the inclusion criteria. The samples were again predominantly middle aged (Sample 1: $M = 43.05$, $SD = 11.87$; Sample 2: $M = 42.48$, $SD = 12.28$) and White (71.0%; 77.1%), but slightly more balanced with respect to gender (57.7% female; 52.5% female).

Materials and Procedure

The procedure for Study 2 was very similar to that of Studies 1a and 1b. Participants first read the same introduction to the construct of verbal ability and rated their own verbal ability. They then read the same introduction to the anagrams task, with participants being randomly assigned to either pre-commit to their prediction of how long they would need to solve two

[4] Exclusion criteria can be found at https://osf.io/sgjwu/wiki/home/.

anagrams (pre-commitment condition) or just think of their prediction (hindsight condition). At this point, orthogonal to the pre-commitment/hindsight manipulation, about half of participants were randomly selected and given two minutes to solve two anagrams, the same as in Studies 1a and 1b (experience condition). The other half did not have the opportunity to solve anagrams (no experience condition) and were instead told: *People tend to underestimate how long it would take to solve anagrams like the ones you just saw. For instance, consider the two anagrams below:* Participants were shown the two anagrams they would have received in the experience condition without their solutions. They then read: *These anagrams are similar in their level of difficulty to the anagrams you saw previously. Past research has shown that people, on average, require around two minutes to solve* ***each*** *anagram.*

As in Studies 1a and 1b, participants in all conditions were then reminded of their initial predictions. Participants in the pre-commitment condition were shown the prediction that they had pre-committed to; participants in the hindsight condition were asked to remember their prediction and record it using the same drop-down menu. Participants then reevaluated their verbal ability. Finally, participants completed the five-item measure of task validity from Study 1. This was again an exploratory measure. All materials and data can be found at https://osf.io/sgjwu/.

Results

Confirmatory Analyses[5]

As in Studies 1a and 1b, belief updating was measured by subtracting participants' initial verbal ability rating from their final verbal ability rating. A 2 (pre-commitment vs. hindsight) x 2 (experience vs. no experience) ANOVA revealed a main effect of pre-commitment, $F(1, 3049) =$

[5] Confirmatory analyses were preregistered at https://osf.io/uwcq3/ (for the first sample) and https://osf.io/ywtrk/ (for the second sample).

13.66, $p < .001$, $p_{augmented} = [.05002, .05003]$, $\eta_p^2 = .003$, a main effect of experience, $F(1, 3049) = 13.95$, $p < .001$, $p_{augmented} = [.050, .05002]$, $\eta_p^2 = .017$, and a suggestive interaction of the pre-commitment and experience manipulations on belief updating, $F(1, 3049) = 4.41$, $p = .036$, $p_{augmented} = [.052, .072]$, $\eta_p^2 = .001$. Breaking apart this interaction, among participants in the experience condition, those who pre-committed to their predictions ($M = -1.85$, $SD = 2.29$) updated their beliefs more than those in the hindsight condition ($M = -1.48$, $SD = 1.91$), $t(1,485) = 3.39$, $p = .001$, $d = .18$, 95% CI [.07, .28]. However, this was not the case in the no experience condition, with those who pre-committed ($M = -1.18$, $SD = 1.85$) and those who did not ($M = -1.11$, $SD = 1.67$) reporting similar belief updating, $t(1,564) = 0.86$, $p = .388$, $d = .04$, 95% CI [-.06, .14]. These effects were not due to differences in initial ratings of verbal ability (range of *M*s: 8.35-8.54, range of *SD*s: 1.78-1.87). Across both the pre-commitment and hindsight conditions, participants in the experience condition updated their beliefs more than those in the no experience condition, pre-commitment condition: $t(1,513) = 6.22$, $p < .001$, $d = .32$, 95% CI [.22, .42]; hindsight condition: $t(1,536) = 4.05$, $p < .001$, $d = .21$, 95% CI [.11, .31]. The effect of experience mirrors patterns from Studies 1a and 1b, in which participants in both conditions tended to lower their ratings of their ability after completing the challenging task.

I also analyzed predicted times to complete the task using a 2 (pre-commitment vs. hindsight) x 2 (experience vs. no experience) ANOVA. There was a main effect of pre-commitment condition, $F(1, 2836) = 62.51$, $p < .001$, $\eta_p^2 = .042$, but neither a reliable main effect of experience, $F(1, 2836) = 1.80$, $p = .179$, $\eta_p^2 = .001$, nor an interaction of pre-commitment and experience, $F(1, 2836) = .44$, $p = .505$, $\eta_p^2 < .001$. Collapsing across experience conditions, those who pre-committed to their predictions ($M = 44.38$, $SD = 34.86$) predicted that they would need

less time to solve the two anagrams compared to those who provided postdictions (M = 60.89, SD = 42.35), $t(2,838) = 11.22$, $p < .001$, $d = .42$, 95% CI [.35, .50].

Exploratory Analyses - Predicted vs. actual performance

The anagram task was again difficult for participants, with only 20.9% of participants who completed the task correctly solving both anagrams within two minutes. Participants in the pre-commitment and hindsight conditions were similarly likely to solve the anagrams (139 in pre-commitment condition; 165 in hindsight condition), $d = .17$, 95% CI [-.06, .40]. Only 8.1% completed both anagrams in as much or less time than they predicted they would need. As in Studies 1a and 1b, participants were more likely to perform in line with their postdictions in the hindsight condition, when they reported their recollection of their prediction after completing the task (N = 80), compared to those in the pre-commitment condition (N = 38), $d = .76$, 95% CI [.37, 1.15].

Discussion

Pre-committing to predictions increased belief updating, but only among participants who actually completed the anagrams task. Those who were simply told that their prediction was likely wrong based on other people's performance showed similar belief updating regardless of whether they pre-committed to a prediction or reported postdictions. This suggests that pre-commitment may only increase belief updating (relative to not pre-committing) when individuals experience the extent to which their predictions were wrong. However, the moderating effect of direct experience on the effect of pre-commitment was relatively small ($\eta_p^2 = .001$) and the inferential evidence for it was only suggestive ($p = .036$, $p_{augmented}$ = [.052, .072]).

Participants were again more likely to have correct predictions when reporting their predictions after the task rather than before it. Without pre-commitment, individuals tend to alter

their predictions to be more in line with their performances, thus increasing their rates of being "correct." This post-hoc adjustment may be one factor contributing to the effect of pre-commitment – participants in the hindsight condition may perceive their predictions as being less incorrect and therefore not feel the need to adjust their beliefs as much.

Study 3 - Manipulating validity of the task

The anagrams task in Studies 1a-2 was designed in a way that may have limited participants' ability to discount its results. The task was always preceded by a description of verbal ability and was described as diagnostic of that ability. Those features may have made it too difficult to dismiss the task as relevant to their beliefs and forced them to consider the evidence. If participants were given more flexibility to interpret the task, however, they may have responded to the larger inconsistencies produced by pre-commitment by discounting the evidence altogether (Bastardi, Uhlmann, & Ross, 2011; Greenwald & Ronis, 1978; Lord, Lepper, & Ross, 1979). I examined this possibility in Study 3.

Methods

Participants

Participants were again recruited from a large online panel. I planned to collect 1,500 participants who passed a series of attention check questions.[6] This sample size was determined by constraints in access to the panel, but would also provide adequate power (97%) to detect effect sizes similar to those of Studies 1a-2 (~d = .20). In total, 1,497 participants were collected that met the inclusion criteria. The sample was predominantly middle aged (M = 46.03, SD = 18.18), White (74.2%), and female (73.0%).

Materials and Procedure

[6] Exclusion criteria can be found at https://osf.io/7u5sb/.

The primary new feature of Study 3 was manipulating whether or not the anagrams task was described as diagnostic of verbal ability. I altered several features of the previous procedure to manipulate perceived diagnosticity.

First, unlike the previous studies, participants did not provide initial ratings of their verbal ability. Instead, participants were simply told:

> *For the next task, you will be asked to solve two anagrams. There will be two strings of letters that can be unscrambled to form words. For instance:*
>
> *ciles (answer = **slice**)*
>
> *terny (answer = **entry**)*
>
> *You will have two minutes to solve these anagrams.*

I removed the initial verbal ability rating and any mention of verbal ability in the task introduction to avoid framing the anagrams task as a measure of verbal ability.

At this point, those in the pre-commitment condition were asked to estimate and record the amount of time they predicted they would need to solve the anagrams, using the same dropdown menu as in previous studies. Unlike in previous studies, those in the hindsight condition were not prompted to make an estimate. This change was meant to further reduce the perception of the anagrams as a test.

Participants were then randomly assigned to either receive information about the diagnosticity of the anagrams task or not. Those in the diagnostic condition read:

> ***More information about this task***
>
> *Researchers have demonstrated that the ability to solve anagrams is indicative of one's overall verbal ability.*
>
> *Verbal ability is an important quality. Verbal ability includes essentials like spelling,*

> *grammar, and reading comprehension. These core skills can combine to predict people's ability on many tasks related to success in school, the workplace, and life in general. Those who can solve anagrams, especially those who can solve them quickly, tend to show higher verbal ability across many different tests. Although brief, tasks like the one you are about to complete can reliably predict who will succeed in domains where verbal ability is important.*

Participants in the control condition simply proceeded to the anagrams task.

The anagrams task was similar to the previous two studies; participants were given two minutes to solve two anagrams with their remaining time presented on the task page. Afterward, all participants, including those in the pre-commitment condition, were asked: *When you were initially told about the anagrams task, before completing it, how long did you think you would need to answer both anagrams?* Participants responded using the same dropdown menu as previous studies. This change allowed me to examine whether those who pre-committed to their predictions accurately recalled their predictions or misremembered their predictions after the task. Then, all participants assessed the validity of the anagrams task, first reading:

> *Verbal ability is an important quality. Verbal ability includes essentials like spelling, grammar, and reading comprehension. These core skills can combine to predict people's ability on many tasks related to success in school, the workplace, and life in general.*
>
> *The anagrams task you just completed was designed to be a test of verbal ability. Based on your knowledge of the task, please tell us how much you agree or disagree with the following statements:*

They then rated their agreement (1 – *Strongly disagree* to 7 – *Strongly agree*) with four statements: *Solving anagrams is related to verbal ability*; *Verbal ability is much more than the ability to solve anagrams* (reverse scored); *Being able to solve anagrams quickly is part of verbal ability*; *Solving two anagrams is too short of a test to provide useful information about verbal ability* (reverse scored). These statements were averaged to create a measure of diagnosticity.

Finally, participants rated their verbal ability, reading:

As previously mentioned, verbal ability is an important quality. Verbal ability includes essentials like spelling, grammar, and reading comprehension. These core skills can combine to predict people's ability on many tasks related to success in school, the workplace, and life in general.

How would you rate your verbal ability compared to other people?

They responded using the same 11-point scale as used in Studies 1a-2. All materials and data can be found at https://osf.io/5me3u/.

Results

Confirmatory Analyses[7]

Unlike in the previous studies, participants only reported their beliefs about their verbal ability after the anagrams task. A 2 (pre-commitment vs. hindsight) x 2 (diagnostic vs. control) ANOVA revealed no main effect of pre-commitment, $F(1, 1491) = 1.23$, $p = .268$, $\eta_p^2 = .0002$, no main effect of diagnosticity, $F(1, 1491) = 2.68$, $p = .102$, $\eta_p^2 = .001$, and no interaction of the pre-commitment and diagnosticity manipulations on verbal ability ratings, $F(1, 1491) = 1.14$, $p = .286$, $\eta_p^2 = .001$. Among those in the diagnostic condition, participants reported similar verbal

[7] Confirmatory analyses were preregistered at https://osf.io/75mzr/.

ability in the pre-commitment (M = 7.23, SD = 2.11) and hindsight conditions (M = 7.40, SD = 2.03), $t(728) = 1.09$, $p = .274$, $d = .08$, 95% CI [-.06, .23]. Among those in the control condition, participants reported similar verbal ability in the pre-commitment (M = 7.22, SD = 2.00) and hindsight conditions (M = 7.16, SD = 2.04), $t(763) = -.40$, $p = .691$, $d = -.03$, 95% CI [-.17, .11]. Finally, participants rated the anagrams task as similarly diagnostic in the diagnostic (M = 3.50, SD = .95) and control conditions (M = 3.47, SD =.82), suggesting that the manipulation was ineffective, $t(1,477) = .48$, $p = .629$, $d = .03$, 95% CI [-.08, .13].

I also examined whether beliefs about verbal ability were predicted by ratings of diagnosticity of the task and the difference between predicted and actual performance on the anagrams task. The latter measure was scored as the difference between participants' predicted time, in seconds, needed to solve the anagrams (provided before the task in the pre-commitment condition and after the task in the hindsight condition) and the amount of time they spent attempting to solve the anagrams. I first constructed a multiple regression model predicting verbal ability ratings from diagnosticity ratings, diagnosticity condition (diagnostic vs. control), and their interaction. This model revealed that verbal ability ratings were related to ratings of diagnosticity, $b = -.22$, $SE = .08$, $p = .005$, $\eta_p^2 = .012$; the effect of condition, $b = .08$, $SE = .43$, $p = .855$, $\eta_p^2 = .001$, and the interaction were not reliable, $b = -.06$, $SE = .12$, $p = .592$, $\eta_p^2 = .0002$. I constructed a second multiple regression model predicting verbal ability ratings, but this time from the difference between predicted and actual performance on the anagrams task, pre-commitment condition (pre-commitment vs. hindsight), and their interaction. This model suggested that verbal ability ratings were related to the difference between predicted and actual performance, $b = -.004$, $SE = .001$, $p = .017$, $\eta_p^2 = .003$; the effect of condition, $b = -.12$, $SE = .13$, $p = .333$, $\eta_p^2 = .0002$, and the interaction were not reliable, $b = .002$, $SE = .002$, $p = .271$, η_p^2

= .001. These results may provide some insight into how individuals are updating their beliefs, suggesting that individuals think about their evaluation of the task and how well they performed on it when assessing their verbal ability.

Replicating the previous studies, I examined the participants' predictions about the amount of time they would need to solve the two anagrams. As in the previous studies, participants in the pre-commitment condition (M = 49.38, SD = 33.12) predicted that they would need less time than those in the hindsight condition (M = 72.12, SD = 42.95), $t(1,387) = 10.78$, $p < .001$, d = .58, 95% CI [.47, .69]. Unlike in the previous studies, participants in the pre-commitment condition were asked to report their predictions a second time, after the anagrams task (in Studies 1a-2, participants were shown their predictions). This second report (M = 65.85, SD = 41.05) was also reliably lower than postdictions provided by the hindsight condition, $t(1,493) = 2.88$, p = .004, d = .15, 95% CI [.05, .25], but also reliably higher than initial predictions in the pre-commitment condition ($M_{difference}$ = 6.28, $SD_{difference}$ = 20.58), $t(601) = -7.49$, $p < .001$, d = -.31, 95% CI [-.39, -.22]. Most participants (75.6%) reported the same prediction that they had previously recorded, and a substantial minority (21.2%) reported a prediction higher than they had actually predicted before completing the task.

Finally, I sought to confirm the exploratory findings from the previous studies that examined predicted vs. actual performance in successfully completing the anagrams task. Only 19.2% of participants correctly solved both anagrams within two minutes. These participants were divided relatively evenly among the pre-commitment and hindsight conditions (128 in pre-commitment condition; 160 in hindsight condition), X^2 (1, N = 288), = 3.56, p = .059, d = .22, 95% CI [-.01, .46]. Only 9.2% completed both anagrams in as much or less time than they predicted they would need. Replicating prior findings, participants were more likely to perform

in line with their postdictions in the hindsight condition, when they reported their recollection of their prediction after completing the task (N = 89), compared to those in the pre-commitment condition (N = 48), X^2 (1, N = 137) = 12.27, p = .0005, d = .62, 95% CI [.27, .98]. This effect holds when using the second report of predictions by participants in the pre-commitment condition, as only a few participants altered their predictions enough to match or exceed their actual performance (57 in pre-commitment condition; 89 in hindsight condition), X^2 (1, N = 146), = 7.01, p = .008, d = .45, 95% CI [.11, .78].

Discussion

In Study 3, I attempted to manipulate the participants' perceptions of the validity of the anagrams task by either describing it as diagnostic of verbal ability or not. However, this manipulation was ineffective at altering assessments of diagnosticity. Correlationally, participants' assessments of the diagnosticity of the task predicted their assessments of their verbal ability, as did the gap between how long they thought they would need to complete the task and their actual time spent on it. This provides some evidence that participants consider their assessments of the task and their performance when assessing their beliefs, but also suggests that my instructions were ineffective at manipulating these assessments.

Unlike in the first two studies, I did not detect the central effect of pre-commitment on beliefs. Self-assessed verbal ability was similar between the pre-commitment and hindsight conditions, and this was not qualified by the diagnosticity manipulation. There are a few possible explanations for this. First, it could be that the effect of pre-commitment lacks methodological generalizability and is dependent on some of the features of the previous study that I changed to accommodate the diagnosticity manipulation. For example, having participants provide an initial rating of their verbal ability, which was omitted in this study, might raise the perceived stakes of

the task; it provides another opportunity for participants to be wrong. These added stakes may be necessary for pre-commitment to have an effect. Second, it is possible that the earlier studies overestimate the true effect of pre-commitment, or that this study underestimates it. Finally, it is possible that the single time point measure of beliefs about verbal ability was less sensitive than the within-subject difference score used in Studies 1a-2. Examining this possibility, the effect of pre-commitment did become slightly weaker when using only the post-task rating of verbal ability in Studies 1a-2 (aggregate d = .20 becomes d = .16). This difference in effect size would lower expected power from 78% (when only considering those in the diagnostic condition) to 59%. Therefore, it was probably unwise to power the study based on the previous effect sizes that used the within-subjects measure.

The results from Study 3 did, however, confirm prior exploratory analyses comparing predicted vs. actual performance in successfully completing the anagrams task. As in the prior studies, participants were more likely to meet or exceed their postdictions (in the hindsight condition) compared to their predictions (in the pre-commitment condition). This supports the idea that participants are more likely to "succeed" on the task when it is possible to retroactively alter their expectations. When given the same opportunity, most participants in the pre-commitment condition did not alter their predictions, but a subset (21%) did. We cannot estimate the proportion of participants in the hindsight condition that were affected by hindsight bias because they did not report their initial predictions. However, the data suggest that the pre-commitment did discourage differences between participants' predictions and postdictions, even though a substantial minority changed their responses at postdiction.

Study 4 - Belief updating or shifting memory?

In Studies 1-3, belief updating was measured as the difference between participants' post-anagrams rating of their verbal ability and their pre-anagrams rating. A straightforward explanation of participants lowering the verbal ability rating throughout the procedure is that they are responding to their performance on the anagrams task and intentionally lowering their assessment as a result. However, it is possible that participants do not realize that they are lowering their ratings. Rather, participants may not remember their initial ratings and believe that they are providing consistent responses throughout. Previous research has shown that people may retroactively shift their recollection of their previous abilities or performances to support a desired narrative (e.g., Conway & Ross, 1984). Misremembering initial verbal ability ratings as having been lower than actually reported could help participants feel like they accurately assessed their abilities. I examine this possibility in Study 4 by both testing participants' memory of their initial verbal ability ratings and by testing their ability to retroactively estimate what their initial rating would have been.

Methods

Participants

Participants were again recruited from a large online panel. I planned to collect 3,000 participants who passed a series of attention check questions.[8] This sample size was determined by constraints in access to the panel, but would also provide adequate power (88%) to detect effect sizes similar to those of Studies 1a-2 (~d = .20) within each new condition (see *Materials and Procedure*, below). In total, 3,083 participants were collected that met the inclusion criteria.

[8] Exclusion criteria can be found at https://osf.io/qp4hr/.

The sample was predominantly middle aged (M = 39.36, SD = 12.47), White (69.8%), and female (67.9%).

Materials and Procedure

As in Studies 1a-2, participants read a description of verbal ability before completing the anagrams task. Participants were randomly assigned to report their prediction for how long they would need to solve two anagrams either before (pre-commitment condition) or after (hindsight condition) completing the task.

Self-ratings of verbal ability were assessed in one of three ways. Roughly one third of participants (randomly assigned) followed the procedure of Studies 1a-2 and reported pre- and post-anagrams assessments of verbal ability (control condition). A second group of participants provided pre- and post-anagrams assessments, but were also asked, after their post-anagrams assessment, to remember and report their pre-anagrams assessment (remember condition). A final group (would condition) did not provide a pre-anagrams assessment and instead, after providing their post-anagrams assessment, were asked: *Before you completed the anagrams task, you read a description of verbal ability. If you had been asked to rate your verbal ability at that point, before having completed the anagrams task, how would you have rated your verbal ability compared to other people?* All verbal ability ratings were made using the scale used in the previous studies.[9]

Results

Confirmatory Analyses[10]

First, I sought to replicate the primary effect from the previous studies, that pre-commitment to predictions leads to increased belief updating. I examined the difference between

[9] Exact wording of study items can be found at https://osf.io/6t4g2/.
[10] Confirmatory results were preregistered at https://osf.io/unqae/.

participants post- and pre-ratings of their verbal ability among those in the control and remember conditions (those in the would condition did not provide pre-ratings). Unlike Studies 1a-2, participants in the pre-commitment condition (M = -1.31, SD = 2.13) updated their beliefs about their verbal ability to a similar degree as those in the hindsight condition (M = -1.24, SD = 1.90), $t(2046) = .77$, $p = .440$, $d = .03$, 95% CI [-.05, .12]. This was not qualified by whether participants were in the control or remember condition, $F(1, 2044) = .18$, $p = .732$, $\eta_p^2 = .0001$.

Next, I examined participants' memories of their pre-anagrams ratings. In aggregate, participants in the remember condition remembered having reported a lower rating of their verbal ability, before completing the anagrams, than they had actually reported ($M_{difference}$ = -.43, $SD_{difference}$ = 1.43), $t(1,068) = -9.79$, $p < .001$, $d = -.30$, 95% CI [-.42, -.18]. Although most participants (70.7%) in the remember condition accurately recalled and reported their pre-anagrams rating, a sizeable number (23.8%) reported a lower rating than they had initially reported. Very few (5.5%) reported a higher rating than they had initially reported. The difference between true pre-anagrams ratings and recalled pre-anagrams ratings did not vary by pre-commitment condition, $t(1,067) = -.69$, $p = .492$, $d = -.04$, 95% CI [-.16, .08]. Interestingly, among those who did report a lower rating, post-anagrams ratings and the difference between remembered and actual pre-ratings were positively correlated, $r(252) = .43$, 95% CI [.33, .53]. That is, participants who reported lower verbal ability ratings during the post-anagrams assessment also tended to have larger gaps between their remembered initial ratings and true initial ratings.

In the would condition, participants were asked to estimate what their pre-anagrams rating of their verbal ability would have been prior to completing the anagrams task. I compared these retrospective estimates of pre-anagrams ratings to the actual pre-anagrams ratings provided

by participants in the remember and control conditions. Overall, these ratings differed by condition, $F(2, 3073) = 138.81$, $p < .001$, $\eta_p^2 = .083$. Participants in the would condition (M = 6.72, SD = 2.67) provided lower ratings of their verbal ability compared to those in the control condition (M = 8.00, SD = 1.90, $p < .001$[11]) and those in the remember condition (M = 8.01, SD = 1.87, $p < .001$). The pre-anagrams ratings in the latter two conditions did not reliably differ from one another ($p = .992$). Unlike reports of memory of pre-ratings (from the remember condition), retrospective estimates of pre-anagrams ratings did reliably differ by pre-commitment condition, $t(1,023) = 2.78$, $p = .005$, $d = .17$, 95% CI [.05, .30]. When asked to estimate what their pre-anagrams rating of their verbal ability *would* have been, participants in the pre-commitment condition estimated lower ratings (M = 6.53, SD = 2.20) than participants in the hindsight condition (M = 6.92, SD = 2.32), $d = .17$, 95% CI [.05, .30]. Overall, these retrospective ratings were slightly higher than "remembered" pre-ratings from participants who lowered their pre-rating in the remember condition (M = 6.33, SD = 2.19), $d = .17$, 95% CI [.04, .31].

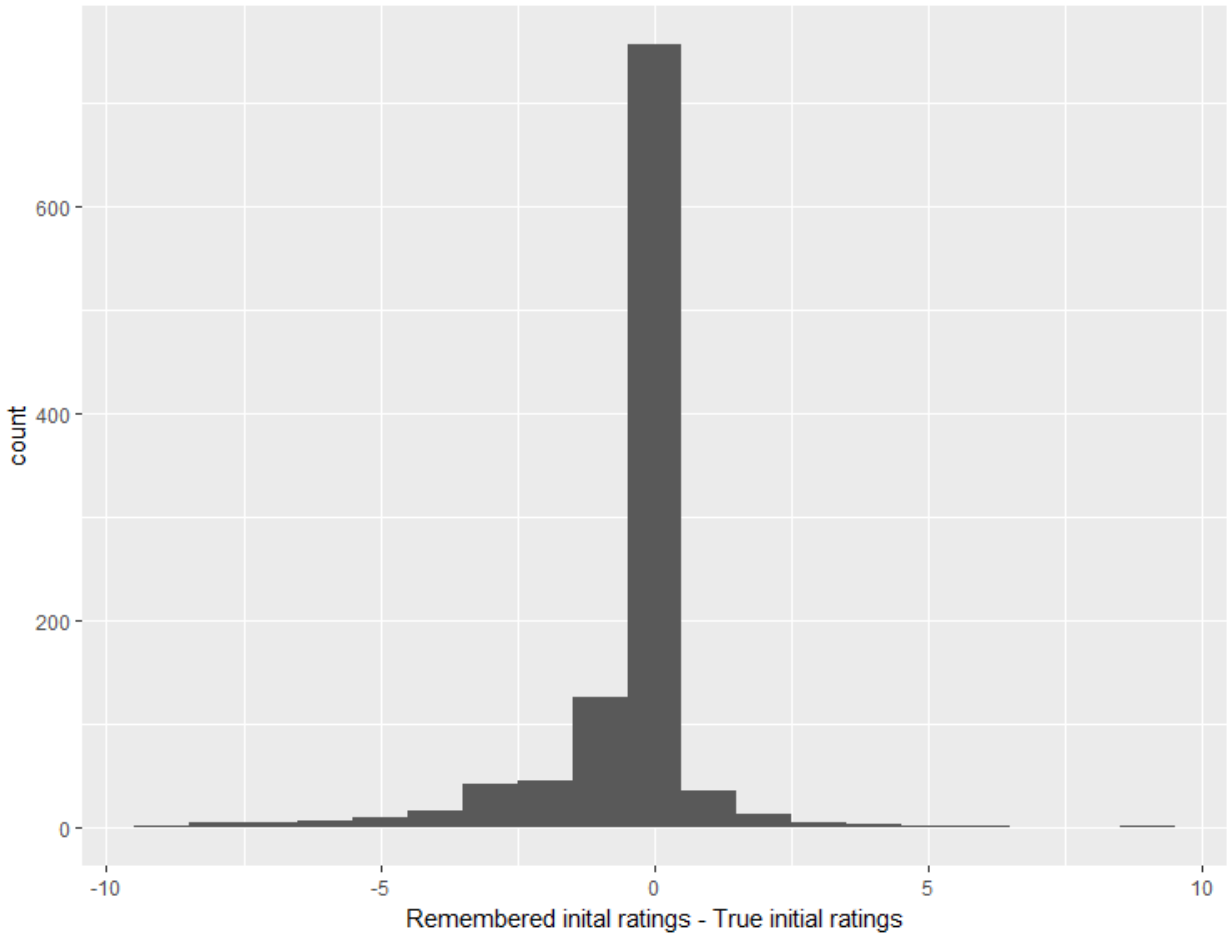
Table 1 - Summary of ability ratings by reporting condition

	True Pre-Rating		Post-Rating		Retrospective Pre-Rating	
Reporting Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	8.00	1.90	6.73	2.37	-	-
Remember	8.01	1.87	6.73	2.36	7.58	2.07
Would	-	-	6.20	2.48	6.72	2.27

[11] Pairwise comparisons corrected for multiple comparisons using the Tukey method.

Note - Retrospective ratings refer to reports of pre-ratings made after the post-rating. In the remember condition, participants were asked to recall their previous rating. Participants in the would condition, who did not make a pre-rating, were asked to estimate how they would have rated their verbal ability before the anagrams task.

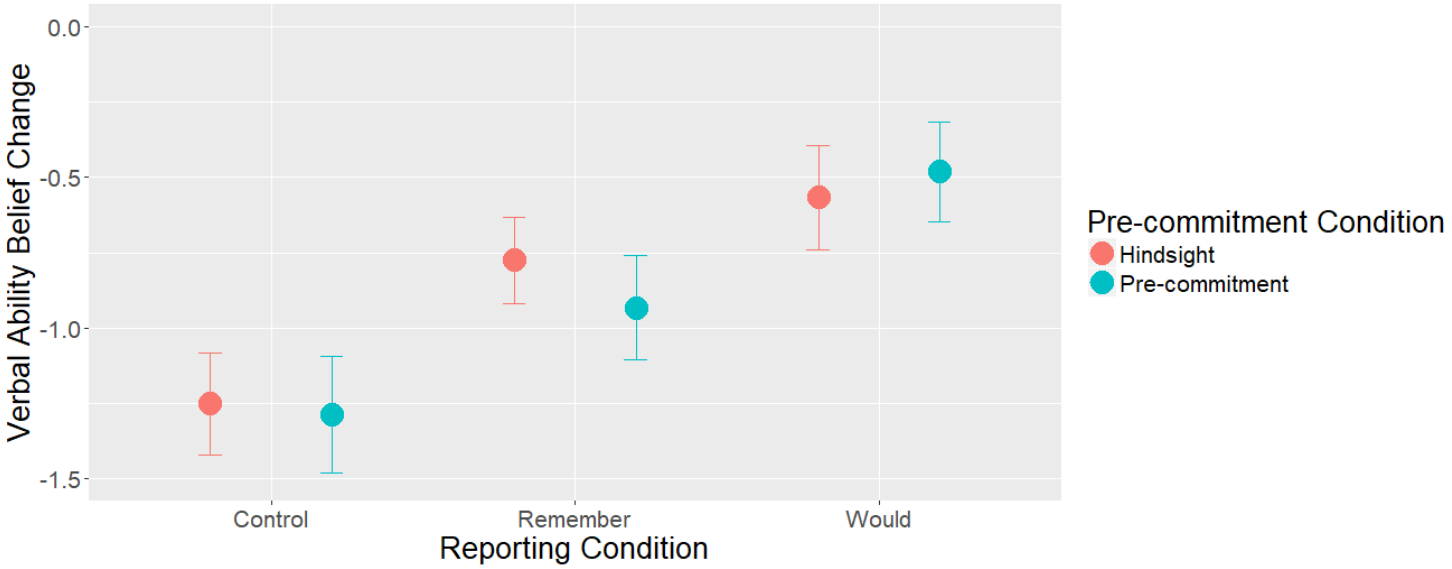
Figure 1 - Distribution of remembered ratings compared to true initial ratings



Note - A score of 0 indicates that a participant accurately recalled and reported their true initial verbal ability rating. A negative value indicates that a participant's remembered initial rating was lower than their true initial rating.

Finally, I examined whether belief updating varied between conditions when using participants' memories of their pre-ratings (in the remember condition) and retrospective estimates of their pre-ratings (in the would condition) to calculate the amount of belief updating. Belief updating did reliably differ between the control, remember, and would conditions, $F(2, 3066) = 16.21$, $p < .001$, $\eta_p^2 = .023$, and this effect was not qualified by pre-commitment condition, $F(2, 3066) = 1.01$, $p = .366$, $\eta_p^2 = .0007$. Participants in the control condition reported the most belief updating (actual pre-anagrams rating - post-anagrams rating, $M = -1.27$, $SD = 2.03$), followed by participants in the remember condition (remembered pre-anagrams rating - post-anagrams rating, $M = -.85$, $SD = 1.88$), followed by participants in the would condition (retrospective estimate of pre-anagrams rating - post-anagrams rating, $M = -.52$, $SD = 1.95$). All pairwise comparisons (Tukey corrected) were significant (all ps $< .001$).

Figure 2 - Belief updating across reporting and pre-commitment conditions



Note: Error bars represent 95% confidence intervals.

Discussion

This study examined whether participants were aware of the extent to which they updated their beliefs about their abilities in light of a challenging task. When asked to remember their initial ratings of the verbal ability, most participants (70.7%) were able to accurately recall their pre-rating. This suggests that most participants are aware of the extent to which they are updating their beliefs. However, a sizeable number (23.8%) remembered having rated their verbal ability lower than they actually had. These percentages were similar to the rates of accurate memory of predictions of performance on the anagrams task observed in Study 4 (75.6% accurate, 21.2% remembered predictions more in line with actual performance).

Participants were less able to retroactively estimate how they would have rated their verbal ability before the anagrams task. These estimates were reliably lower than the actual pre-anagrams ratings provided by participants in the remember and control conditions. Furthermore, retrospective estimates varied by pre-commitment condition, with participants in the pre-commitment condition estimating that they *would* have rated their verbal ability lower than those in the hindsight condition. These results suggest that, in light of completing a difficult task, participants believe they would have been less confident in their abilities than those who actually rated their abilities pre-task. This effect is stronger when participants are required to pre-commit to predictions about how they would perform on a test of those abilities.

Finally, this study failed to replicate the primary finding from Studies 1a-2, being that pre-commitment to predictions increases belief updating. This result, along with the results of Study 3, does, at most, cast doubt on the initial findings and, at least, suggests that the effect may not be as large previously thought. Study 5 will provide more information about the reliability of this effect, while also examining a possible mechanism.

Study 5 - Examining the prediction-results gap as the mechanism for pre-commitment

In the present paradigm, I provide an opportunity for belief change by presenting an ostensible measure of verbal ability and then providing a "failure" experience on that measure. Asking participants to make a pre-commitment to how they believe they will perform on the measure changes beliefs more than not obtaining a pre-commitment. Why? I believe that the mechanism for this is straightforward: Pre-commitment preserves the diagnosticity of predictions. That is, the explicit act of pre-commitment provides a concrete prediction that is difficult to avoid when confronted with actual performance. For instance, an individual who predicts that they will need 60 seconds to solve the anagrams, and completes the task in 55 seconds, may feel like they were successful on the task. Their performance affirmed their self-belief, and it is obvious because they made explicit what they thought would happen before doing the behavior. If that person needs 90 seconds to complete the task, however, they may consider their performance as challenging to their self-belief. Because they made it explicit, they have accountability to their earlier claim, reducing their ability to rationalize the outcome. As such, they should update their beliefs in response to the new evidence. Moreover, like natural Bayesians, if the discrepancy between the prediction and actual performance is larger, they should update their beliefs even more. In sum, I hypothesize that individuals use the magnitude of the gap between predictions and outcomes to update their beliefs about their verbal ability.

Said another way, pre-commitment may have its impact by reducing the likelihood of employing hindsight bias to dismiss disconfirming evidence. Hindsight bias leads individuals to shrink the gap between predictions and results by helping them to reconstruct what they would have predicted to be more in line with what actually occurred. This reduces the discrepancy between "predictions" and performance thus requiring less updating of existing beliefs. In Study

5, I tested whether the magnitude of the difference between predictions and outcomes mediated the effect of pre-commitment on updating beliefs.

Exploratory Mediation Analyses - Studies 1a and 1b

I investigated the proposed mechanism using the data from Studies 1a and 1b. The prediction-results gap for each participant was calculated as the difference between their prediction (in the pre-commitment condition) or postdiction (in the hindsight condition) and actual amount of time that they spent on the anagrams task. Participants in the pre-commitment condition had larger prediction-results gaps (M = -44.39, SD = 44.83) compared to participants in the hindsight condition (M = -23.01, SD = 44.01), d = -.48, 95% CI [-.55, -.41]. Furthermore, across both conditions, participants' prediction-results gaps were correlated with change in beliefs about verbal ability, $r(3127)$ = -.21, 95% CI [-.24, -.17]. That is, participants with larger gaps between their predicted and actual performance on the task tended to update their beliefs more than those with smaller prediction-results gaps. In support of the proposed mediator, there was an indirect effect of the prediction-results gap on belief updating (1000 iterations bootstrapped estimate of indirect effect = -.18, 95% CI [-.22, -.14]). The effect of pre-commitment on belief updating was accounted for, in part, by restricting participants' ability to alter their predictions to be more in line with their actual performance.

These exploratory results were generated from a measurement of mediation design (e.g., Baron & Kenny, 1986; Spencer, Zanna, & Fong, 2005). The causal claim of this design assumes that the mediator covaries with the outcome measure, that the mediator occurs before the outcome measure, and that the mediator is unconfounded (Cook, Campbell, & Shadish, 2002). While the previous studies satisfy the first two assumptions, the design does not sufficiently rule out other possible explanations for the detection of this mediation effect. It could be that the pre-

commitment manipulation independently leads to larger prediction-results gaps *and* greater belief updating with both effects being caused by an unmeasured mediator.

To address this limitation, I attempted to manipulate the proposed mediator in Study 5 (e.g., Pirlott, & MacKinnon, 2016). Half of participants completed the same procedure as Studies 1a and 1b. The other half completed the same procedure, except they were not told how long they would have to complete the anagrams task nor were they shown a timer when completing the task. This should make it more difficult for participants to recognize the discrepancy between their prediction and their results. If the effect of pre-commitment on updating beliefs is mediated by the prediction-results gap, then the indirect effect should be reduced when participants have less information highlighting the magnitude of the gap. However, if the effect of pre-commitment on updating beliefs and the proposed mediator are independent of one another, the indirect effect should be similar regardless of procedure.

Methods

Participants

Participants were recruited from a large online panel. I planned to collect 3,000 participants who passed a series of attention check questions.[12] This sample size was determined by constraints in access to the panel, but would also provide adequate power (97%) to detect effect sizes similar to those of Studies 1a-2 (~d = .20) within each timer condition (see *Materials and Procedure*, below). In total, 3,172 participants were collected that met the inclusion criteria. The sample was predominantly middle aged (M = 39.03, SD = 12.73), White (67.0%), and female (63.2%).

Materials and Procedure

[12] Exclusion criteria can be found at https://osf.io/mzh6u/.

The materials and procedure for this study were very similar to Studies 1a and 1b. First, participants rated their verbal ability before estimating how long they would need to solve two anagrams. Those in the pre-commitment condition reported this prediction; those in the hindsight condition were simply be asked to think about their prediction. Then, all participants were given two minutes to solve two new anagrams.

Half of participants completed the anagrams task as presented in Studies 1a and 1b. They were told prior to the task that they would have two minutes to complete both anagrams. They were also shown a timer while completing the anagrams so that they knew how much time they had spent on the task (time known condition). The other half were not told how long they would have to complete the task and were not shown a timer during the task (time unknown condition). If they had not completed the anagrams after two minutes, they were told that, in the interest of time, they must move on to the next portion of the study.

Afterwards, participants were reminded of their prediction, either by being shown their reported prediction (in the pre-commitment condition) or by asking them to recall and report it (in the hindsight condition). All participants then re-rated their verbal ability. Finally, all participants were asked to estimate how long they had spent on the anagrams task (in seconds).

Results

Confirmatory Analyses[13]

First, I examined the effectiveness of the timer manipulation. If removing information about allotted time and the timer during the anagrams task made it difficult for participants to estimate how long they had spent on the task, participants in the time known condition should have more accurate estimates (a smaller absolute difference between estimated time and actual

[13] Confirmatory results were preregistered at https://osf.io/jfrz3/.

time spent on the task) than those in the time unknown condition. This was the case, with participants in the time known condition reporting more accurate assessments of how long they had spent on the anagrams task (M = 20.05, SD = 24.13) compared to those in the time unknown condition (M = 30.47.86, SD = 27.61), $t(3135) = 11.23$, $p < .001$, $d = .40$, 95% CI [.33, .47].[14]

The primary focus of this study was experimentally testing the difference between predicted and actual performance on the anagrams task as a mediator of the effect of pre-commitment on belief updating. I examined this question in two ways. First, I examined the pathways of a traditional mediation model within each timer condition separately. I then tested the timer manipulation as a moderator of the indirect effect.

The first path in the mediation model is the direct effect of the pre-commitment manipulation on belief updating. This path was not statistically reliable in either condition, in the time known condition: $t(1546) = 1.24$, $p = .215$, $d = .06$, 95% CI [-.04, .16]; in the time unknown condition: $t(1608) = .20$, $p = .845$, $d = .01$, 95% CI [-.09, .11]. The lack of an effect of pre-commitment in the time known condition was contrary to predictions. However, indirect effects can still be detected in the absence of a significant direct effect (Rucker, Preacher, Tormala, & Petty, 2011). Therefore, I proceeded with testing the rest of the mediation model.

The second path in the mediation model is the effect of the pre-commitment manipulation on the difference between predicted and actual performance. In the time known condition, participants in the pre-commitment condition reported larger discrepancies between their predicted and actual time spent on the anagrams (M = -25.37, SD = 49.76) compared to those in

[14] One concern with the timer manipulation is that the presence of the timer may be threatening to participants and, on its own, cause more belief updating. To examine this possible side effect of the timer, I compared participants who were in the hindsight condition. Based on the proposed model, these participants should have similar amounts of updating, regardless of whether their time was known or unknown. This was indeed the case, with participants in the time known condition (M = -1.36, SD = 1.96) reporting similar belief updating compared to those in the time unknown condition (M = -1.35, SD = 1.92), $d = .01$.

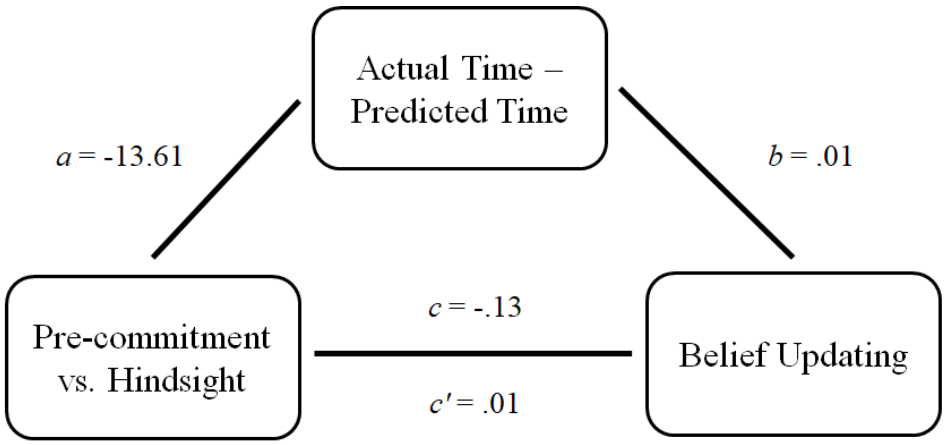
the hindsight condition (M = -11.76, SD = 43.18), t(1537) = -5.74, p < .001, d = -.29, 95% CI [-.39, -.19]. Interestingly, this path was also statistically reliable in the time unknown condition, but in the opposite direction (pre-commitment condition: M = -19.74, SD = 49.39; hindsight condition: M = -26.67, SD = 45.02), t(1594) = 2.92, p = .004, d = .15, 95% CI [.05,.24]. It is possible that this latter effect was due to participants trying to retroactively alter their predictions to fit their actual performance, but being inaccurate in doing so because they lacked the timer on the task. Supporting this possibility, while the predictions reported prior to the anagrams task were similar in the time known and time unknown conditions (d = -.08, 95% CI [-.17, .02]), postdictions were higher in the time known (M = 62.29, SD = 40.59) condition than in the time unknown condition, (M = 51.67, SD = 38.08), d = .27, 95% CI [.17, .37].

Finally, the last path in the mediation model consists of the relation between the prediction-results gap and belief updating. This relation was reliable in both timer conditions, in the time known condition: r(1530) = -.25, p < .001, 95% CI [-.30, -.20]; in the time unknown condition: t(1592) = -.13, p < .001, 95% CI [-.18, -.09].

Next, I estimated the indirect effect of the prediction-results gap on the direct effect of pre-commitment on belief updating. The indirect effect was reliably different from zero in the time known condition, indirect effect = -.15, 95% CI [-.21, -.09][15], F(2, 1556) = 51.40, p < .001. The indirect effect was also reliably different from zero in the time unknown condition but in the opposite direction and smaller in magnitude, indirect effect = .04, 95% CI [.01, .08], F(2, 1610) = 15.18, p < .001. Finally, the timer manipulation did reliably moderate the proposed mediation effect, difference of indirect effects = .19, 95% CI [.13, .25], p < .001.

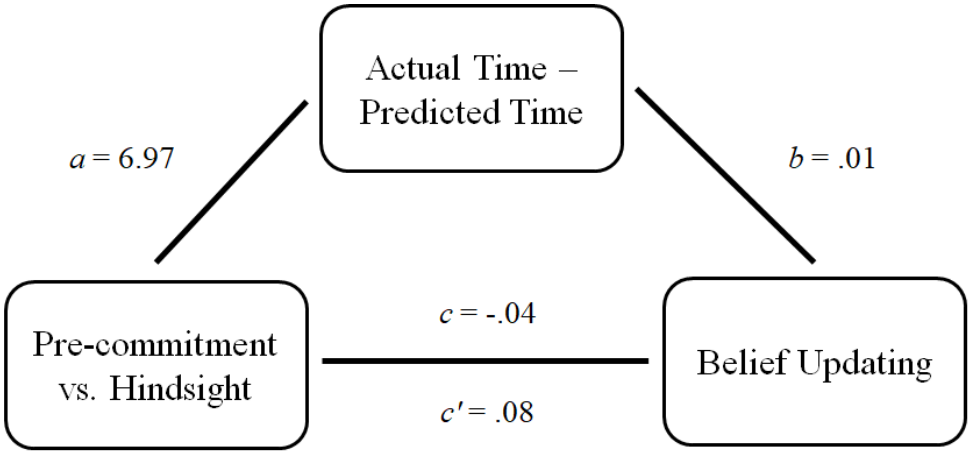
[15] All mediation effects estimated with 1000 bootstrap resamples.

Figure 3a - Mediation pathways in time known condition



Indirect Effect = -.15, 95% CI = [-.21, -.09]

Figure 3b - Mediation pathways in time unknown condition



Indirect Effect = .04, 95% CI = [.01, .07]

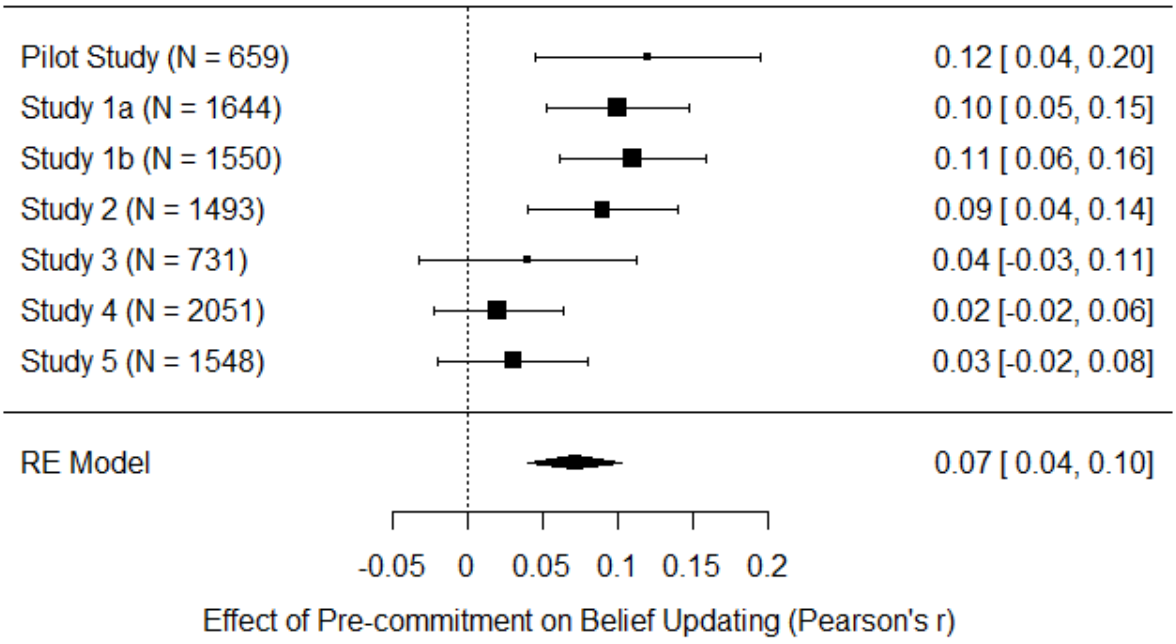
Discussion

In Study 5, I sought to replicate and further test a possible mechanism for the effect of pre-commitment on belief updating. While the direct effect was not statistically reliable in the predicted condition, the indirect effect of the difference between predicted and actual performance was successfully replicated. Furthermore, an experimental suppression of the mediator did reduce and even reverse the direction of the indirect effect. These results provide evidence in favor of the proposed mechanism. Participants seem to use the discrepancy between how they thought they would perform on the anagrams task and how they actually performed to update their ratings of their verbal ability. The effect of pre-commitment on that updating is due, in part, to preventing individuals from retroactively altering their predictions to more closely match their actual performance. However, the lack of a direct effect of pre-commitment in the current study does encourage some caution when interpreting these findings. Although, indirect effects can be present without a significant direct effect (Rucker, Preacher, Tormala, & Petty, 2011), the absence of the direct effect does run counter to the proposed model.

Internal Meta-Analysis

While directionally consistent, the direct effect of pre-commitment on updating beliefs has been statistically inconsistent through these studies. To provide an aggregate estimate of this effect, I conducted an internal meta-analysis. I included data from studies 1a-5 as well as from an initial pilot study. From each of these studies, I selected the comparison of conditions that most resembled the two-celled design from Studies 1a and 1b (see Appendix A for list of conditions). In total, this included seven effect size estimates from 9,676 participants. In aggregate, the effect of pre-commitment on belief updating was small, but reliably different from zero, $r = .07$, 95% CI [.03, .11], $z = 4.46$, $p < .001$.

Figure 4 - Internal meta-analysis of effect of pre-commitment on updating beliefs



Study 6 - Applying pre-commitment to prevent motivated interpretation of evidence

So far, I have demonstrated that pre-commitment to predictions can reduce hindsight bias and, to a small degree, produce more belief updating. Pre-commitment may be similarly useful in other instances where motivated interpretations of evidence can shape beliefs. Confirmation bias may be one such instance (Nickerson, 1998). If individuals trust new evidence more when it supports their existing views, they may be less likely to update their beliefs when encountering mixed evidence. To examine this possibility, I implemented pre-commitment as an extension to a past study that demonstrated confirmation bias in the context of interpreting scientific evidence (Bastardi, Uhlmann, & Ross, 2011).

Methods

Participants

Participants were recruited from a large online panel. I planned to collect 1,500 participants who passed a series of attention check questions.[16] This sample size was determined by constraints in access to the panel. The final sample consisted of 1,514 participants. The sample was predominantly middle aged (M = 51.27, SD = 11.66), White (85.5%), and female (64.3%).

Materials and Procedure

The study was presented as an investigation of how individuals process new information. Participants were first introduced to the topic of that new information: the effect of day care vs. home care on children's' development.[17] Participants read a brief description of the debate between the two forms of care before providing an initial rating of which type of care they believed was best for children's development (1 - *day care far superior*, 5 - *no significant differences*, 9 - *home care far superior*).

Next, participants read about two studies that compared the effects of day care and home care on children. One randomly assigned children to either receive home care or day care. The other examined statistically matched samples of children in day care and home care. Both studies examined a number of outcomes for the children, such as their intelligence, emotional attachment to their parents, and social behaviors with other children. One study always concluded that home care was superior to day care; the other study always concluded that day care was superior to home care. Which study supported which outcome was counterbalanced between participants.

---

[16] Exclusion criteria can be found at https://osf.io/bv6uy/.
[17] See https://osf.io/n83ks/ for all materials.

All participants rated how convincing they would find the results of each study (1 - *extremely unconvincing/invalid*, 4 - *neither convincing nor unconvincing*, 7 - *extremely convincing/valid*). I created a difference score between the ratings of each study, such that higher numbers indicated that the study that supported home care was more convincing than the study that supported day care. All participants also rated whether they found the methods of either study more valid than the other (1 - *random assignment much more valid*, 5 - *both equally valid*, 9 - *statistical control much more valid*). Responses on this measure were transformed such that higher numbers indicated that the study that supported home care had more valid methods. I averaged these two items (the difference score and the rating of methods validity) as an index of relative study quality.

Participants were randomly assigned to rate study quality either with or without knowledge of the results of each study. In the pre-commitment condition, participants read the description of each study's methodology without the results included. They then completed the measures of study quality. Afterwards, they reread the descriptions of the studies, this time with their results included, and were reminded of their assessments of study quality. In the hindsight condition, participants read the results-included versions of the studies only, and then completed the measures of study quality.

Finally, participants rerated which type of care they considered best for children's development. Participants then reported if they had or planned to have children and which type of care (day care or home care) they primarily used or planned to primarily use for their children.

Results

Confirmatory Analyses[18]

[18] Confirmatory results were preregistered at: https://osf.io/bv6uy/.

The primary focus of this study was examining whether pre-commitment altered the relations between pre-ratings of home care vs. day care, ratings of study quality, and post-ratings of home care vs. day care. I first examined the relation between ratings of study quality and pre-ratings. I constructed a linear multiple regression model predicting ratings of study quality from pre-ratings, pre-commitment condition (pre-commitment vs. hindsight), and the interaction of pre-ratings and pre-commitment condition. This interaction term was significant, b = -.16, SE = .03, $p < .001$, $\eta_p^2 = .024$. Decomposing this interaction, pre-ratings were positively correlated with ratings of study quality in the hindsight condition, $r(738) = .27$, 95% CI [.20, .33], $p < .001$, whereas the two were uncorrelated in the pre-commitment condition, $r(751) = -.04$, 95% CI [-.11, .03], $p = .292$. Among those who initially favored home care, participants in the hindsight condition found the study that supported home care more convincing (M =.44, SD = 1.06) than participants in the pre-commitment condition (M = 0.04, SD = 1.12), d = .37, 95% CI [.23, .51]. Similarly, among those who initially favored day care, participants in the hindsight condition found the study that supported day care more convincing (M = -.12, SD = 1.03) than participants in the pre-commitment condition (M = 0.05, SD = 1.11), d = -.25, 95% CI [-.45, -.06]. This replicated the findings of Bastardi, Uhlmann, and Ross (2011) - when participants knew the results of the studies while rating their quality, they tended to rate the study that supported their initial view more positively than the study that did not support their initial view, regardless of which study supported their position. And, extending those findings, when the results were not known, there was no evidence of confirmation bias - study quality assessments were unrelated to initial beliefs.

Next, I examined the relation between post-ratings and ratings of study quality. I constructed a linear multiple regression model predicting post-ratings from ratings of study

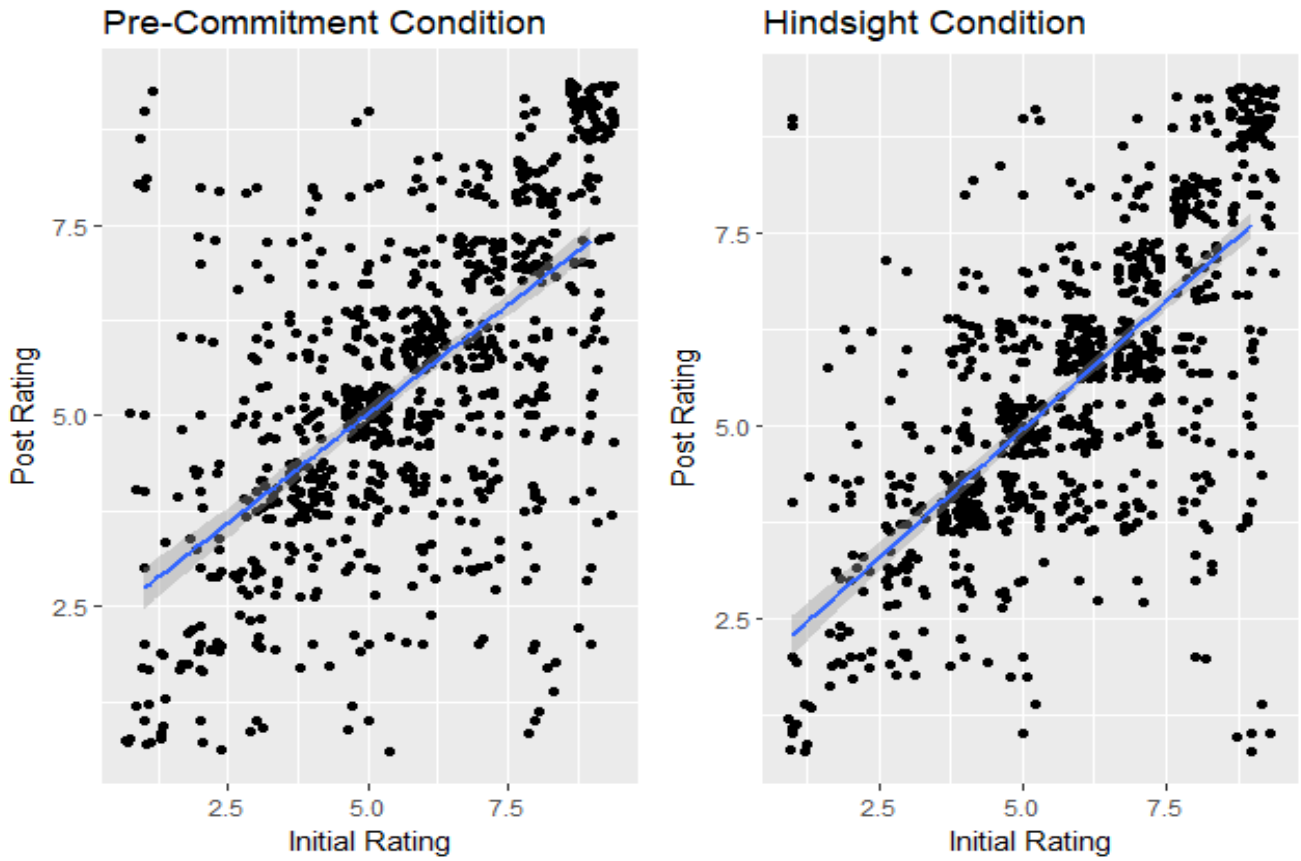
quality, pre-commitment condition (pre-commitment vs. hindsight), and the interaction of ratings of study quality and pre-commitment condition. This interaction term was also significant, b = -.59, SE = .09, $p < .001$, $\eta_p^2 = .027$. Decomposing this interaction, ratings of study quality were positively correlated with post-ratings in the hindsight condition, $r(740) = .32$, 95% CI [.25, .38], $p < .001$, whereas the two were uncorrelated in the pre-commitment condition, $r(752) = -.004$, 95% CI [-.08, .07], $p = .910$. The lack of correlation in the pre-commitment condition may be due to how individuals tended to rate the studies without knowing their results. Participants in the pre-commitment condition rated the two studies as similarly convincing (M = .04, SD = 1.08), with 36.8% of participants rating the studies as equally convincing.

Finally, I examined the relation between pre-ratings and post-ratings. I constructed a linear multiple regression model predicting post-ratings from pre-ratings, pre-commitment condition (pre-commitment vs. hindsight), and the interaction of pre-ratings and pre-commitment condition. This interaction term was suggestive, b = -.10, SE = .04, p = .007, $\eta_p^2 = .005$. While the relation between pre- and post-ratings was strong and positive in both conditions, the relation was slightly stronger in the hindsight condition, $r(463) = .71$, 95% CI [.67, .74], $p < .001$, compared to in the pre-commitment condition, $r(757) = .62$, 95% CI [.57, .66], $p < .001$.

To further examine the relations between pre- and post-ratings, I calculated a difference score between the two ratings. Overall, both participants who initially favored home care ($M_{difference}$ = -.93, $SD_{difference}$ = 1.54) and participants who initially favored day care ($M_{difference}$ = .78, $SD_{difference}$ = 1.50) tended to become more moderate in their views (that is, less in favor of their initial position) after reading about the two studies. Across both initial positions, those in the pre-commitment condition (pro-home care: $M_{difference}$ = -1.03, $SD_{difference}$ = 1.61; pro-day care: $M_{difference}$ = .97, $SD_{difference}$ = 1.69) tended to moderate their views more than those in the

hindsight condition (pro-home care: $M_{difference}$ = -.84, $SD_{difference}$ = 1.47; pro-day care: $M_{difference}$ = .58, $SD_{difference}$ = 1.25), pro-home care: d = -.12, 95% CI [-.26, .01]; pro-day care: d = .26, 95% CI [.06, .45]. This suggests that the difference in correlations between pre- and post-ratings is due to participants in the pre-commitment condition changing their views on childcare more than those in the hindsight condition.

Figure 5 - Relation between pre-ratings and post-ratings by condition



Discussion

Initial opinions on childcare shaped assessments of scientific studies comparing home care to day care. However, if those studies were evaluated based on their methodology and not their results, initial opinions no longer predicted assessments of study quality. Eliminating this confirmation bias seemed to have consequences for belief updating as well. Overall, participants

tended to report more moderate views on childcare after reading the studies, perhaps reflecting the fact that they had just read mixed evidence in favor of their initial belief. This updating was stronger, though, for those who pre-committed to their evaluations of the studies. These results provide initial evidence that pre-commitment can be used to reduce confirmation bias in assessments of new evidence and increase belief updating.

General Discussion

Seven studies (total N = 14,979) examined the effects of pre-commitment on updating beliefs. These studies required individuals to pre-commit to predictions about their upcoming performance on a task (Studies 1a-5) or pre-commit to evaluations of scientific studies before learning their results (Study 6). In both cases, pre-commitment reduced the likelihood that individuals would retroactively alter some part of their judgment of new evidence to be more in line with the observed results. Reducing the likelihood of such alterations led to larger gaps between individuals' initial views and/or predictions and the new results. These larger gaps seemed to encourage more belief updating.

The pre-commitment manipulation highlighted the differences between predictions and postdictions. In line with previous work (Campbell & Tesser, 1983; Fischhoff, & Beyth, 1974; Hom & Ciaramitaro, 2001), postdictions (predictions reported after results are known) were more "accurate" than predictions produced by pre-commitment (Studies 1a-5). In the context of the anagrams paradigm, participants' postdictions, relative to predictions, were more likely to meet or exceed actual performance on the task. Similarly, in Study 6, participants' initial beliefs were correlated with their assessments of the quality of new scientific evidence, as long as they made those assessments with knowledge of the studies' results. Without knowledge of results, initial views were uncorrelated with assessments of study quality. In both paradigms, these

retroactive shifts in predictions or judgments may have served to bolster individuals' assessments of their abilities (in the case of the anagrams paradigm) or reinforce their existing beliefs (in the case of evaluating scientific evidence). Overall, these results demonstrate that, without pre-commitment, individuals are more likely to retroactively alter their predictions or judgments of evidence to more closely fit results. Pre-commitment, therefore, preserves the diagnosticity of such evaluations - it insures that they are not biased by knowledge of results.

Interestingly, the act of pre-commitment seems sufficient to discourage most people from engaging in hindsight bias. In Studies 4 and 5, participants were given the opportunity to retroactively alter their prediction for how long they would need to complete the anagrams task (Study 4) or alter their initial verbal ability rating (Study 5). In each case, most (~70%) did not alter their response. In contrast, participants who did not pre-commit seemed either unable or unwilling to simulate what their initial predictions about performance or initial ability ratings would have been. Pre-commitment may lead to stronger encoding of predictions, reducing memory errors that can increase hindsight bias (Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988) or increase feelings of accountability (Lerner & Tetlock, 1999). At least in the brief paradigms employed by these studies, it seems that pre-commitment is sufficient to encourage (mostly) accurate reporting of previous responses.

In aggregate, those who pre-committed to predictions updated their beliefs more than those who did not. This effect was mediated by increased gaps between predictions and results (Study 5). That is, pre-commitment led individuals to have predictions that more strongly mismatched their results relative to those who did not pre-commit. Individuals seemed to use this gap as information when updating their beliefs. These results are consistent with previous models of belief updating (e.g., Anderson, 1971; Kruglanski, 1990; Kruglanski, Dechesne, Orehek, &

Pierro, 2009) in that individuals were sensitive to new information and updated their beliefs accordingly. Pre-commitment enhances this process by reducing post-hoc adjustments based on observed results.

Participants' sensitivity to the magnitude of the discrepancy between their predictions and results also suggests a kind of Bayesian updating on the part of participants. Although some have found that people tend to not use Bayesian reasoning (e.g., base rate neglect, Kahneman & Tversky, 1972), others have found that people are capable of Bayesian updating when priors are expressed as frequencies rather than probabilities (Gigerenzer & Hoffrage, 1995). The discrepancy between predictions and results, in these studies, may have been intuitive enough for participants to engage in Bayesian thinking and updating. It may be the case that similar effects would not be found in paradigms where the discrepancy between predictions and results was more difficult to comprehend.

Finally, Study 6 examined the effect of pre-commitment in the context of confirmation bias (Nickerson, 1998; Wason, 1966; 1968). Confirmation bias, like hindsight bias, requires knowledge of results when assessing a new piece of evidence. Pre-committing to an evaluation of a new piece of evidence before knowing its results (e.g., evaluating a study's methodology before knowing its outcomes) should lead individuals to produce less biased assessments of new evidence. Study 6 provided preliminary evidence that pre-commitment may be an effective intervention to confirmation bias. As in Studies 1a-5, the effect of pre-commitment on belief updating, when reducing confirmation bias, was small (d = -.14, .29). Further studies are needed to validate the effect of pre-commitment on belief updating in the context of reducing confirmation bias.

Understanding the effect of pre-commitment

The effect of pre-commitment on belief updating was reliable overall but weak (r = .07, 95% CI [.03, .11]). There are several possible factors that could be relevant for understanding the observed effect size. Most obviously, the effect of pre-commitment on belief updating could be relatively weak in general, regardless of the particular features of the paradigms I used. Results from the proposed mechanism may support this possibility. Although postdictions more closely matched participants' performance on the anagrams task compared to predictions, postdictions tended to not fully match results. That is, participants in the hindsight condition did not always report a postdiction that was consistent with how long they spent on the anagrams task. There might be a limit to which people are comfortable retroactively altering their predictions. This could put a ceiling on the extent to which pre-committing to predictions, relative to reporting postdictions, will produce greater belief updating - an intervention can only be as strong as the behavior it seeks to reduce.

Conversely, features of the present paradigm may underestimate the potential of pre-commitment on updating beliefs. For example, the anagrams paradigm was very brief, comprised of trying to solve two anagrams in two minutes. This provides little disconfirming evidence on its own. Anagrams also represent one of many possible tests of verbal ability. These features may make the evidence produced by the test not particularly compelling for participants. However, evidence against this possibility is the fact that the brief intervention was effective in reducing participants beliefs in their verbal ability overall (range: d = -.63 to -.86). Nevertheless, it is possible that more impactful interventions would increase the impact of pre-commitment on subsequent beliefs. The effect of pre-commitment will be better understood with further

investigations using different tests of different beliefs as well as with multiple and/or more in-depth exposures to new evidence.

Finally, it is important to note that the current studies examined pre-commitment in a narrow range of conditions. The studies predominantly made use of the anagrams paradigm – the extent to which the effect of pre-commitment generalizes to other contexts is largely unknown. Study 6 provides some evidence that pre-commitment can affect belief updating in the context of other biases and other materials, but more studies will be needed to establish the generalizability and boundary conditions of this finding. Also, although the studies sampled from United States adults, broadly, there may be some groups for which these effects are stronger or weaker. For instance, individuals with more intellectual humility (e.g., "superforecasters" in geopolitical prediction competitions, Mellers et al., 2014) may be more influenced by new evidence and show stronger effects of pre-commitment (Leary et al., 2017).

Pre-commitment and belief change over time

Pre-commitment reduces the impact of biases in interpreting new information and can therefore increase the likelihood of having to confront disconfirming evidence. This increased likelihood, over time, may change the trajectory of information integration processes and resulting beliefs. Information Integration Theory (Anderson, 1971) provides a straightforward example of this change. Without pre-commitment, biases such as hindsight bias and confirmation bias would likely produce interpretations of new evidence that support existing beliefs. These new confirming pieces of information would be added to existing beliefs, strengthening them over time. Consistent use of pre-commitment could limit these biases, leading to much more mixed evidence for existing beliefs. The addition of these less-biased interpretations of new

evidence could produce weaker beliefs or, to the extent that a belief is often wrong, perhaps a reversal of a belief.

In addition to reducing bias in interpretations, consistent pre-commitment may also affect a person's motivation to continue testing their beliefs. From the standpoint of Lay Epistemology (Kruglanski, 1990; Kruglanski, Dechesne, Orehek, & Pierro, 2009), once a person is satisfied with the demonstrated accuracy of a belief (e.g., their belief seems to adequately explain some part of the world), they stop seeking to test that belief. Encountering more disconfirming information, which seems to be a consequence of pre-commitment, should delay the process of reaching a satisfactory belief and encourage further testing of beliefs (Kruglanski & Freund, 1983; Mayseless & Kruglanski, 1987). If individuals are able to access valid tests of their beliefs, pre-commitment should expedite the process of calibrating beliefs with reality. However, the overall process of developing a satisfactory belief may take longer as people are required to reconcile more conflicting results.

But what if we're right?

The current studies leveraged situations where individuals were likely to encounter new evidence that challenged their beliefs. Pre-commitment was meant to be a tool to facilitate belief updating in the face of that challenging evidence. However, it is important to also understand the effect pre-commitment would have in situations where individuals have correct beliefs (that is beliefs supported by new evidence). It is possible that in these situations, pre-commitment to predictions may have no discernible effect on belief updating. Individuals are often unaware of their biases or reliance on heuristics (Pronin, Gilovich, & Ross, 2004; Pronin, & Kugler, 2007; Pronin, Lin, & Ross, 2002). Therefore, they might think that predictions are just as valid and informative as postdictions when compared to new evidence.

Conversely, pre-commitment may strengthen beliefs more when confirmed by new evidence. Much like a bank shot in basketball is more impressive when called by the shooter ahead of time, people may put more weight in correct predictions when they know that they committed to those predictions before seeing the results. To a certain extent, this would be a rational way to update beliefs; individuals should have more faith in assessments of evidence that are shielded from bias. However, if correct predictions occur by chance (which is bound to happen), pre-commitment may exacerbate incorrect beliefs. Correct pre-committed predictions, and the added confidence they afford, may encourage individuals to stop testing their beliefs earlier than they otherwise would have, increasing the leverage of a given test on resulting beliefs (Kruglanski & Freund, 1983; Mayseless & Kruglanski, 1987).

Public vs. private commitment

Feelings of accountability can reduce biases and reliance on heuristics (Lerner & Tetlock, 1999). When individuals know that they are accountable to others for their judgments and decisions, they tend to rely on numeric anchoring less (Kruglanski & Freund, 1983), be less influenced by ordering effects (Kruglanski & Freund, 1983; Tetlock, 1983), and be less influenced by sunk cost concerns (Simonson & Nye, 1992). Pre-commitment may similarly make individuals feel accountable. Within the current studies, participants were aware that the researcher(s) had recorded their pre-committed predictions and judgments. In a way, this made participants' pre-commitments public (to at least some group of individuals). This might have made them feel more beholden to those responses, and that they therefore needed to update their beliefs more. Conversely, if individuals knew that their pre-commitments would not be known to anyone but themselves, they may discount those pre-commitments and update their beliefs less.

Future work is needed to investigate whether pre-commitment instills feelings of accountability and if those feelings are necessary for increased updating of beliefs.

Applicability to scientific research

Scientific researchers sometimes incorporate pre-commitment into their research workflow. One practice, referred to as preregistration, entails creating a public, time-stamped record of hypotheses and/or analysis plans before observing new data (Nosek, Ebersole, DeHaven, & Mellor, 2018). Preregistration is meant to shield the scientist from biases when analyzing and interpreting data and to increase the diagnosticity of predictions and statistical tests (most prominently, null hypothesis significance testing, de Groot, 2014; Simmons, Nelson, & Simonsohn, 2011; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievitet, 2012). Another example of pre-commitment in research is the Registered Reports model of peer review and publishing (Chambers, 2013; Nosek & Lakens, 2014). In this model, a research project's rationale, methods, and planned analyses are peer reviewed before the results are known. This process is meant to ensure that publication decisions are made on the basis of research quality, not outcomes (Hergovich, Schott, & Burger, 2010; Mahoney, 1977), and reduce judgment biases of authors and reviewers based on the favorability or unfavorability of the results for their beliefs.

The current studies suggest that these practices may have some benefits for bias reduction in the creation of scientific evidence. Pre-commitment to predictions prevented individuals from retroactively altering their predictions to match their observed data. This reduced the number of supposedly "correct" predictions within the paradigm. Similarly, preregistration in research should reduce the likelihood that researchers alter their hypotheses or their analysis plans to fit results observed in the data. In Study 6, pre-commitment reduced the relation between initial

beliefs and assessments of the quality of studies that tested those beliefs. The Registered Reports process may likewise reduce the impact of researchers' prior beliefs on their evaluations of new research.

These debiasing effects give reason to believe that pre-commitment, if broadly instituted in research, would similarly lead to increased belief updating about scientific findings and theories. Preregistration and Registered Reports place the evaluative focus on the process of reaching results rather than on the results themselves. Therefore, if tested hypotheses or theories are correct, these methods should lead to more confirming evidence. If tested hypotheses or theories are incorrect, however, these methods should lead to more disconfirming evidence. This would produce a less biased account of the scientific literature and, as a result, hopefully more accurate models of the world. This is in contrast to traditional models of scientific inquiry and dissemination which lack pre-commitment and seem to bias in favor of producing confirming, rather than disconfirming, information (Franco, Malhotra, & Simonovits, 2014; Greenwald, 1975; Kaplan & Irvin, 2015; Sterling, 1959). Early evidence suggests that Registered Reports reduce these biases, with the results of Registered Reports supporting the null hypothesis 55-66% of the time, compared to 5-20% of traditional articles supporting the null (Allen & Mehler, 2018).

Conversely, there are reasons to suspect that pre-commitment may have less of an impact on scientific beliefs. Science is meant to operate in a cumulative fashion. Scientific beliefs should, ideally, be based on a larger body of knowledge rather than isolated studies. Pre-commitment, at least as I have operationalized it, operates on the level of single tests or studies. Pre-commitment may increase the likelihood of encountering disconfirming evidence in research, but its impact, as generated from a single study, may be limited on a more multiply-

determined scientific belief. Greater adoption of preregistration and Registered Reports may, as demonstrated in these studies, increase the amount of disconfirming results in a given area of research, shaping the resulting beliefs of researchers. However, it is unclear how much disconfirming evidence would be needed to meaningfully shift beliefs about scientific findings nor is it clear how widespread pre-commitment would need to be in order to produce that amount of disconfirming evidence.

Another important factor for the effect of pre-commitment on scientific belief updating is the extent to which scientists agree on the interpretation of results. In the anagrams paradigm, determining whether predictions were accurate or inaccurate was fairly straightforward -- was the predicted time needed to solve the anagrams more or less than what was actually observed? In scientific research, judging whether a prediction was accurate is often more difficult. Even in potentially the most straightforward case of scientific verification, direct replication, there is not broad consensus as to what constitutes a "successful" or "failed" test of a phenomenon or theory (Maxwell, Lau, & Howard, 2015; Open Science Collaboration, 2015; Simonsohn, 2015). As such, pre-commitment may have less of an impact on the resulting beliefs of the scientific community. If pre-commitment primarily impacts the individual that is testing their beliefs, but has little impact on those observing that test, its effect on scientific consensus will be small.

Conclusion

Beliefs help people navigate their world. New evidence can lead people to alter their beliefs so that they better reflect reality. However, certain biases can shape how individuals interpret that new evidence and potentially limit its impact on belief updating. The current studies suggest that pre-commitment to predictions and evaluations of evidence can reduce the

impact of those biases, making it more likely that individuals will have to confront evidence that conflicts with their beliefs. This leads to increased belief updating.

Finding places where we are wrong is useful. Disconfirming evidence provides an opportunity to revise our beliefs so that they will hopefully become more accurate. Pre-commitment expedites this process by showing us the full extent to which our beliefs are wrong. This provides a better opportunity to alter our beliefs so that we can make better predictions and choices in the future.

References

Allen, C. P. G., & Mehler, D. M. A. (2018, October 17). Open Science challenges, benefits and tips in early career and beyond. https://doi.org/10.31234/osf.io/3czyt

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *78*(3), 171.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science, 22*(6), 731.

Bodgan, R. J., (1986). The importance of belief. In R. J. Bodgan (Ed.), *Belief: Form, content, and function*. Oxford, England: Clarendon Press.

Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, *51*(4), 605-620.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752.

Chaiken, S. (1987). The heuristic model of persuasion. In *Social influence: The Ontario Symposium* (Vol. 5, pp. 3-39).

Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex, 49*(3), 609-610.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*(1), 147-168.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472-517.

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808.

Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology*, *47*(4), 738.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

de Groot, A. D. (2014). The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta Psychologica*, *148*, 188-194.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568.

Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, *29*(9), 1120-1132.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*(1), 53.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.

Fischhoff B., & Beyth R. (1975). 'I knew it would happen'--Remembered probabilities of once-future things *Organizational Behavior and Human Performance, 13*(1), 1–16.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502-1505.

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition*, *30*(6), 652-668.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*(4), 684.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*(1), 1.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*(7), 603.

Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review*, *85*(1), 53.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*(6), 533-538.

Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, *29*(3), 188-209.

Hom, H. L., & Ciaramitaro, M. (2001). GTIDHNIHS: I knew-it-all-along. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(5), 493-507.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454.

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PloS One*, *10*(8), e0132382.

Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 51-60.

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, *1*(3), 181-197.

Kruglanski, A. W., Dechesne, M., Orehek, E., & Pierro, A. (2009). Three decades of lay epistemics: The why, how, and who of knowledge formation. *European Review of Social Psychology*, *20*(1), 146-191.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, *19*(5), 448-468.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480.

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., ... & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, *43*(6), 793-813.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098.

Mackie, D. M., Worth, L. T., & Asuncion, A. G. (1990). Processing of persuasive in-group messages. *Journal of Personality and Social Psychology*, *58*(5), 812.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research,1*(2), 161-175.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist*, *70*(6), 487.

Mayseless, O., & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes*, *39*(2), 162-183.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., ... & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*(5), 1106-1115.

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1399.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606.

Nosek, B. A. & Lakens D. (2014) Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*, 137–141.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of experimental social psychology, 66*, 29-38.

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*(4), 565-578.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411-426.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, *5*(6), 359-371.

Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science*, *9*(3), 293-304.

Sanna, L. J., & Schwarz, N. (2006). Metacognitive experiences and human judgment: The case of hindsight bias and its debiasing. *Current Directions in Psychological Science*, *15*(4), 172-176.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559-569.

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*(3), 416-446.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*(285), 30-34.

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 285-292.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632-638.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology I*, Harmondsworth, Middlesex, England: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273-281.

Appendix A - Conditions included in internal meta-analysis

Selection Criteria

The internal meta-analysis sought to estimate an aggregate effect size for the effect of pre-commitment on updating beliefs. I included conditions from each study that most resembled the initial paradigm used to demonstrate the focal effect. The following sets of conditions, and effect sizes and sample sizes, were included in the meta-analysis:

Pilot Study - An initial pilot used a design very similar to Studies 1a and 1b. However, I tried two different introductions to the construct of verbal ability (one that included more detail about the construct and its importance and a much briefer description). I included the two conditions (hindsight vs. pre-commitment) that accompanied the longer description of verbal ability, as this was the design used in subsequent studies (N = 659, r = .12).

Study 1a - This study had two experimental conditions (hindsight vs. pre-commitment) and both were included in the meta-analysis (N = 1644, r = .10).

Study 1b - This study had two experimental conditions (hindsight vs. pre-commitment) and both were included in the meta-analysis (N = 1550, r = .11).

Study 2 - From this study, I included the two conditions (hindsight vs. pre-commitment) that accompanied the experience condition (participants who actually completed the anagrams task, N = 1493, r = .09).

Study 3 - From this study, I included the two conditions (hindsight vs. pre-commitment) that accompanied the diagnostic condition (participants who read about verbal ability and the diagnosticity of the anagrams task before completing it, N = 731, r = .08).

Study 4 - From this study, I included the two conditions (hindsight vs. pre-commitment) that accompanied both the control condition and the remember condition. This choice was made so

that all participants would have a pre- and post-anagrams task verbal ability rating (the would condition did not provide a pre-rating, $N = 2051$, $r = .02$).

Study 5 - From this study, I included the two conditions (hindsight vs. pre-commitment) that accompanied the time known condition ($N = 1548$, $r = .03$).