

Harder Than You Think: How Outside Assistance Leads to Overconfidence



Matthew Fisher¹ and Daniel M. Oppenheimer²

¹Marketing Department, Cox School of Business, Southern Methodist University, and

²Department of Social and Decision Sciences, Carnegie Mellon University

Psychological Science
2021, Vol. 32(4) 598–610
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797620975779
www.psychologicalscience.org/PS



Abstract

Cognitive ability consists not only of one's internal competence but also of the augmentation offered by the outside world. How much of our cognitive success is due to our own abilities, and how much is due to external support? Can we accurately draw that distinction? Here, we explored when and why people are unaware of their reliance on outside assistance. Across eight experiments ($N = 2,440$ participants recruited from Amazon Mechanical Turk), people showed improved metacognitive calibration when assistance occurred after a delay or required active choice. Furthermore, these findings apply across a wide range of cognitive tasks, including semantic memory (Experiments 1a and 1b), episodic memory (Experiments 2a and 2b), and problem solving (Experiments 3a–3d). These experiments offer important insights into how we understand our own abilities when we rely on outside help.

Keywords

metacognition, illusions of knowledge, augmented cognition, open data, open materials, preregistered

Received 3/31/20; Revision accepted 9/22/20

The human mind faces inherent limitations. However, many constraints can be mitigated through off-loading (Risko & Gilbert, 2016). People adeptly outsource mental activities to compensate for limited internal abilities. Thus, cognitive activity can be viewed as the combination of both internal and external operations (Clark & Chalmers, 1998). As a result, when trying to understand a cognitive system, one must consider more than just what is “in the head.”

Throughout history, humans have developed ways to outsource mental activities. Ancient Peruvians used intricately knotted cords as memory aids for recalling important events (Tylor, 1870), and ancient Romans trained slaves to remember legal information so that they could be called on in a public debate (Nestojko et al., 2013; Schönplflug & Esser, 1995). In everyday life, people routinely externalize memory by relying on tools such as notebooks, calendars, and lists (Block & Morwitz, 1999; Harris, 1980; Intons-Peterson & Fournier, 1986). Indeed, people sometimes exhibit a bias toward the use of external rather than internal resources, even when it is costly (Gilbert et al., 2020). These strategies improve memory for new information by reducing the interference caused by internally

stored information (Henkel, 2014; Runge et al., 2019; Storm & Stone, 2015). The seamlessness of outside assistance could make tracking the source of information quite difficult.

Metacognitive Errors

Overconfidence pervades many types of self-assessment (Dunning, 2012; Wilson, 2009). Across a variety of dimensions, people consider themselves better than average (Alicke & Govorun, 2005). However, in many domains, people remain oblivious to the shallowness of their own knowledge. They exhibit an *illusion of explanatory depth*, overestimating how much detail they can provide about how everyday objects work (Rozenblit & Keil, 2002), arguments on controversial topics (Fernbach et al., 2013; Fisher & Keil, 2014), and topics within their area of expertise (Fisher & Keil, 2016). People do not naturally interrogate their own

Corresponding Author:

Matthew Fisher, Southern Methodist University, Cox School of Business, Marketing Department
E-mail: matthewfisher@smu.edu

depth of knowledge; they appraise how easily surface knowledge comes to mind (e.g., “Do I know what a toilet is?”) but do not spontaneously query how well they understand the details (e.g., “How does the float ball regulate the intake of water from the inlet valve?”; Alter et al., 2010). Put another way, only when people have attempted to recall information do they recognize the difficulty involved in retrieval.

Confidence in one’s own knowledge is often driven by the subjective experience related to that particular information (Koriat, 1997; Mueller et al., 2016). Consequently, metacognition relies on inferential and heuristic processes instead of direct access, leading learners to engage in suboptimal study habits (Bjork et al., 2013). One key metacognitive heuristic, *retrieval fluency*, produces overconfidence when information feels easy to process (Alter & Oppenheimer, 2009; Frank & Kuhlmann, 2017; Whittlesea, 1993; Yan et al., 2016). In sum, metacognition is prone to systematic error, whereby people consider themselves to be more competent than is warranted. These studies have focused on how well people can assess their own abilities but have not considered how the extended nature of the mind could support these illusions.

Metacognition and the Outsourced Mind

Previous research has focused on how people decide to off-load memory. When people are not confident in their own abilities, they off-load memory to external tools (Boldt & Gilbert, 2019; Dunn & Risko, 2016). For example, people will use a computer to save the most difficult words in a memory task (Hu et al., 2019). Further, people off-load memory when they judge the external aid to be reliable (Weis & Wiese, 2019). Here, we asked a related but distinct question: When off-loading has already occurred, do external sources distort metacognitive assessments of one’s own abilities? Building on the finding that the Internet induces the illusion of knowledge (Fisher et al., 2015), we aimed to uncover the principles underlying metacognitive judgments involving distributed cognition. We argue that external assistance that can be seamlessly integrated with one’s own abilities makes it difficult to accurately assess the extent to which one’s performance should be attributed to one’s own ability.

When people complete a task without help, they are largely aware of the amount of cognitive effort that they applied (see Thomson & Oppenheimer, 2020, for a discussion of the nature of cognitive effort). When receiving cognitive support, people are still aware of the amount of work they have done but have no metacognitive access to the work that was done externally or how much effort it would have taken to have done that work themselves. For example, people have no

Statement of Relevance

When people outsource cognition by using technology to support their thinking (e.g., looking up facts on Google), they often do not realize how heavily they are relying on that outside assistance. This leads them to overestimate their own abilities (e.g., thinking that they know more than they really do). Here, we conducted interventions for reducing this bias: time delays and active choice. In doing so, we gained insight into why the bias exists in the first place. These findings contribute to our understanding of metacognition (i.e., our awareness of our own abilities), human–machine interaction, and the promise and pitfalls of augmented cognition, and they can help predict and prevent errors of overconfidence.

sense of the difficulty or confusion associated with a math problem when a calculator does the work. The process of entering the numbers and reading the output is highly fluent and easy, and the internal cues for signaling difficulty are absent. As with the illusion of explanatory depth, only by interrogating internal cognitive systems (see Alter et al., 2010) can people get a sense of how they would have done in the absence of external augmentation.

If the aforementioned logic is correct, then improved calibration occurs when people experience the effort of completing the task on their own prior to receiving help. Just as “desirable difficulties” improve learning (Bjork & Bjork, 2011), interrupting the fluency of receiving external assistance may improve metacognitive accuracy. This suggests two underlying reasons for why people fail to properly calibrate estimates of their own abilities. The first of these is immediacy. When people receive assistance without delay, they do not have a chance to query their own abilities and thus never have access to internal metacognitive cues (see Koriat, 1997). However, if people must wait before receiving assistance, they will often naturally attempt the task on their own and realize the task’s difficulty.

The second reason is a lack of deliberate choice. In cases in which external cognition is recruited by default, people never need to evaluate their own competence and rarely consider what they are able to do without help. Thus, for tools used by default, people may be more inclined to underestimate how much of the cognitive work is being done by the assistive technology. People’s estimates of their own abilities should be better calibrated when they actively choose to use the technology.

We aimed to test our model of the underlying reasons for metacognitive miscalibration through the

design of interventions to overcome it. We tested these two mechanisms across three distinct domains: semantic memory (Experiment 1), episodic memory (Experiment 2), and problem solving (Experiment 3). Preregistration forms for five of the eight experiments, deidentified raw data files, and analytic syntax are available on OSF at <https://osf.io/ruxmk>. All dependent variables and conditions are reported for all experiments.

Experiment 1a

To explore the effects of delay on the metacognition of extended cognition, we tested a previously validated cognitive domain: retrieval from semantic memory in the form of trivia questions (Fisher & Oppenheimer, in press).

Method

Participants. Two hundred ninety-eight participants from the United States (158 males, 140 females; mean age = 37.85 years, $SD = 21.70$) completed the experiment online through Amazon Mechanical Turk via the TurkPrime platform (Buhrmester et al., 2018; Litman et al., 2017). Sample size was determined so that there was at least an 80% chance of detecting the estimated effect size (Cohen's $d = 0.37$) given previous research using a similar paradigm (Fisher & Oppenheimer, in press). Because of quality concerns on online platforms (Chmielewski & Kucker, 2020), we did not allow participants who failed to correctly answer two initial attention-check questions to continue to the rest of the experiment.¹

Materials, design, and procedure. To begin the experiment, participants were instructed, "Your task will be to correctly answer as many questions as you can in 90 seconds. You are not allowed to use any outside help—please answer all questions on your own." Participants then viewed eight trivia questions while a timer counted down from 1 min 30 s to zero. The questions were taken from the trivia website Sporcle.com and were of intermediate difficulty. Sporcle's data indicated that the percentage of correct responses for these questions ranged from 32.7% ("Which alcoholic spirit gets its primary flavor from juniper berries?" answer: gin) to 77.6% ("Which religion worships Shiva, Devi, Vishnu, Ganesha, and Surya?" answer: Hinduism; see Section S1 in the Supplemental Material for the full stimuli set). Questions were selected such that participants could provide a plausible guess even if they did not know the correct answer. Additionally, they were told, "Once an answer is submitted it cannot be changed. To skip an item, click the arrow below the word." Participants then viewed the eight trivia questions presented in a randomized order.

Each participant was randomly assigned to the no-help, the help, or the delay condition. The no-help condition received no assistance, but participants in the help condition saw the first letter of the correct answer displayed beneath each question. In the delay condition, the initial instructions included the following text: "For each question a hint will appear after 7 seconds. You do not have to wait for the hint—you can answer as soon as you think you know the answer." As participants in the delay condition completed the trivia questions, the text "Hint appears in:" appeared below each question. Next to this text, a timer counted down from 7 s to zero. Once the timer reached zero, the first letter of the correct answer appeared. All participants were given 1 min 30 s to complete the trivia portion of the experiment. After completing the eight trivia questions, all participants were told how many of the questions they had answered correctly. Finally, all participants were asked, "If you were to answer another set of similar questions without any outside help or hints, what percentage would you answer correctly?"²

Results

Participants in the delay condition answered as many trivia questions correctly ($M = 62.63\%$, $SD = 23.49\%$) as those in the help condition ($M = 59.44\%$, $SD = 23.36\%$), $\beta = 0.11$, $SE = 0.12$, $p = .36$, and significantly more than those in the no-help condition ($M = 33.59\%$, $SD = 22.56\%$), $\beta = 1.08$, $SE = 0.12$, $p < .001$.³ Despite equivalent scores, participants in the delay condition provided lower estimates of their future performance ($M = 51.42\%$, $SD = 28.69\%$) relative to those in the help condition ($M = 58.54\%$, $SD = 25.83\%$), $\beta = -0.28$, $SE = 0.13$, $p = .03$ (see Fig. 1). The delay condition predicted higher performance relative to the no-help condition ($M = 32.72\%$, $SD = 25.33\%$), $\beta = 0.62$, $SE = 0.13$, $p < .001$ (see Fig. 1).

As expected, participants' estimates of future performance were much higher when they received help than when they did not; they appeared to attribute their success to internal abilities rather than the hints that they had received. This finding replicated previous work using a similar paradigm (Fisher & Oppenheimer, in press).

The novel element of the present experiment was the addition of the delay condition. Introducing a delay partially mitigated but did not entirely eliminate the difference in self-assessed ability. This is consistent with the theory that introducing a delay allows participants to query their memories before the metacognitive cues are compromised by a hint that makes retrieval easier, giving a more accurate sense of what performance would be like without assistance. In Experiment 1b, we

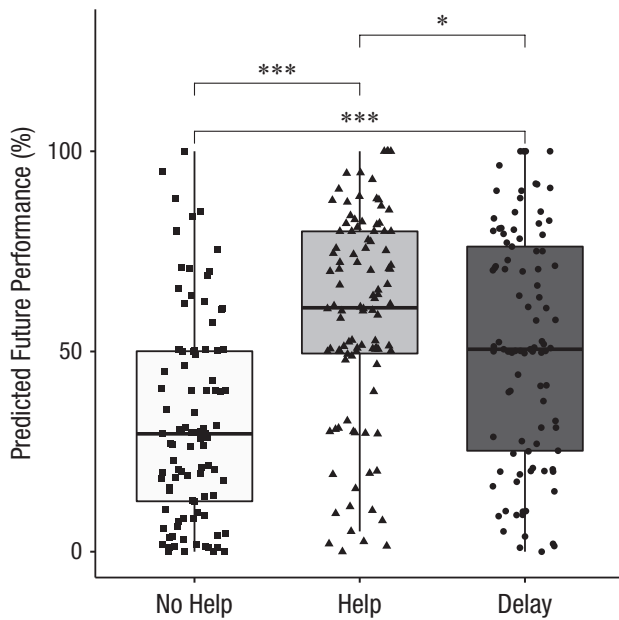


Fig. 1. Predicted future performance (percentage correct) by condition in Experiment 1a. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions ($*p < .05$, $***p < .001$).

tested the second intervention drawn from this logic—requiring participants to actively choose to receive assistance.

Experiment 1b

In Experiment 1b, we adopted the same paradigm as in Experiment 1a, except that instead of receiving help after a delay, participants needed to actively click a button to receive help.

Method

Participants. Two hundred ninety-six participants from the United States (181 males, 115 females; mean age = 36.27 years, $SD = 11.87$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. The instructions and materials were the same as in Experiment 1a. Each participant was randomly assigned to one of three conditions: no help, help, or button. In the no-help condition, participants answered the eight trivia questions without any hints. In the help condition, the first letter of the correct answer appeared below each question. In the button

condition, a button labeled “HINT” appeared beneath each question. Participants could click the button and immediately view the first letter of the correct answer. In this paradigm, it is possible for participants to not use the available information, so we tracked how many times participants in the button condition actually clicked the button. After completing the trivia portion of the experiment, all participants were asked, “If you were to answer another set of similar questions without any outside help or hints, what percentage would you answer correctly?”

Even though participants were instructed not to use outside help, it was possible for them to quickly look up the answers to the trivia questions. To detect possible cheating, we used the TaskMaster tool (Permut et al., 2019) to track how long participants navigated away from our experiment. On the basis of our preregistered exclusion criteria, we removed 38 participants from the analysis for having clicked away from the trivia questions for more than 15 s.

Results

Participants in the button condition correctly answered as many questions ($M = 50.41\%$, $SD = 23.76\%$) as those in the help condition ($M = 54.51\%$, $SD = 20.64\%$), $\beta = -0.20$, $SE = 0.14$, $p = .14$, and more than those in the no-help condition ($M = 31.41\%$, $SD = 18.66\%$), $\beta = 0.80$, $SE = 0.14$, $p < .001$. Participants in the button condition predicted significantly lower future performance ($M = 46.38\%$, $SD = 25.57\%$) than those in the help condition ($M = 58.68\%$, $SD = 23.06\%$), $\beta = -0.57$, $SE = 0.14$, $p < .001$, and better future performance than those in the no-help condition ($M = 34.92\%$, $SD = 21.24\%$), $\beta = 0.44$, $SE = 0.14$, $p = .002$ (see Fig. 2). Seventy-seven percent of participants in the button condition clicked to view at least half of the hints. The results remained significant when participants who did not click were excluded from the analysis.

As in Experiment 1a, participants’ estimates of future performance were significantly higher when they received help than when they did not. Forcing participants to engage with the questions without a hint before receiving assistance (by requiring them to push a button before the hint was provided) mitigated but did not eliminate this tendency. This result, in conjunction with the findings of Experiment 1a, is consistent with the argument that people do not naturally attribute the feeling of fluency to the assistance they are receiving. In the absence of feelings of difficulty, participants think the task is easy; it is only when they are forced to bear the full weight of the cognitive processing required (even for a short period) that they realize the true difficulty of the task.

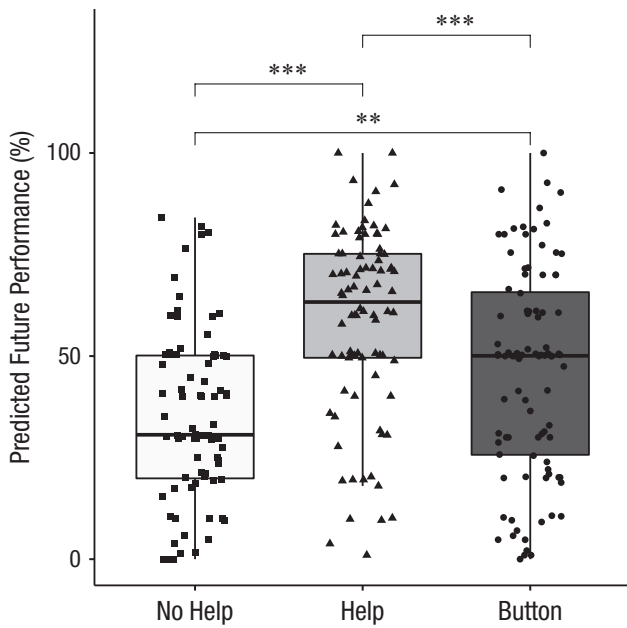


Fig. 2. Predicted future performance (percentage correct) by condition in Experiment 1b. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions (** $p < .01$, *** $p < .001$).

In this experiment, we explored a single aspect of memory—retrieval from semantic memory. Memory outsourcing occurs across a broader array of tasks, and it remains to be seen whether the principles of extended metacognition explored above generalize beyond semantic retrieval. Consequently, in the next set of experiments, we aimed to extend the findings to episodic memory.

Experiment 2a

In Experiment 2a, we tested episodic memory by asking participants to memorize and recall a list of words. Some participants were provided hints, and crucially, some participants were provided hints only after a delay. Participants then predicted how well they would be able to remember future lists of words if they did not have hints.

Method

Participants. Four hundred ninety-nine participants from the United States (260 males, 239 females; mean age = 36.88 years, $SD = 11.98$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. Participants were instructed “to recall as many words as you can. Fifteen

words will be presented on the screen for a total of 20 seconds. Memorize as many as possible in the order they are presented.” Participants then viewed the following randomly selected 15 five-letter words: *pride, doubt, ranch, quote, tread, track, deter, swear, award, ideal, smart, ferry, debut, drink, stall*. After 20 s had elapsed, the survey automatically advanced to the next page. Participants were next instructed to type each word that they remembered, in order, in the blank boxes provided.

After viewing the list of 15 words for 20 s, participants recalled as many items as they could by submitting their answers one at a time. Only answers that appeared on the original list and were submitted in the appropriate order were counted as correct. Each participant was randomly assigned to one of four conditions: no help, help, delay help, or delay only. The no-help condition received no assistance. The help condition viewed the first three letters of the correct answer for each of the 15 words. In the delay-help condition, the first three letters of the correct answer appeared after 7 s. As in Experiment 1a, a timer appeared on the screen that counted down until the hint appeared. In the initial instructions, participants were also notified that the hint would be appearing after a delay. Additionally, to ensure that delays did not, ipso facto, produce lower metacognitive estimations (e.g., by making the task harder because memory traces decay), we included a delay-only condition. In this condition, participants did not receive any hints but could not submit each answer until 7 s had elapsed. Participants in all conditions were given a total of 4 min 20 s to recall the list of words. After the recall portion of the experiment, all participants estimated how well they would perform on another memory test without any outside assistance.

Results

Participants in the delay-help condition remembered as many words ($M = 55.87\%$, $SD = 25.98\%$) as those in the help condition ($M = 61.61\%$, $SD = 22.49\%$), $\beta = -0.18$, $SE = 0.11$, $p = .09$, and more than those in the no-help condition ($M = 26.88\%$, $SD = 25.54\%$), $\beta = 0.96$, $SE = 0.11$, $p < .001$, or delay-only condition ($M = 25.46\%$, $SD = 28.65\%$), $\beta = 0.99$, $SE = 0.11$, $p < .001$. There was no significant difference in accuracy between the delay-only condition and the no-help condition, $p = .79$. Participants' predictions in the delay-help condition ($M = 40.96\%$, $SD = 28.04\%$) were consistent with the previous results: They predicted lower future performance than participants in the help condition ($M = 48.30\%$, $SD = 27.25\%$), $\beta = -0.26$, $SE = 0.11$, $p = .02$, and higher performance than those in the no-help condition ($M = 26.39\%$, $SD = 26.62\%$), $\beta = 0.52$, $SE = 0.11$, $p < .001$, and delay-only condition ($M = 23.11\%$, $SD = 24.93\%$), $\beta = 0.58$, $SE = 0.12$, $p < .001$ (see Fig. 3).

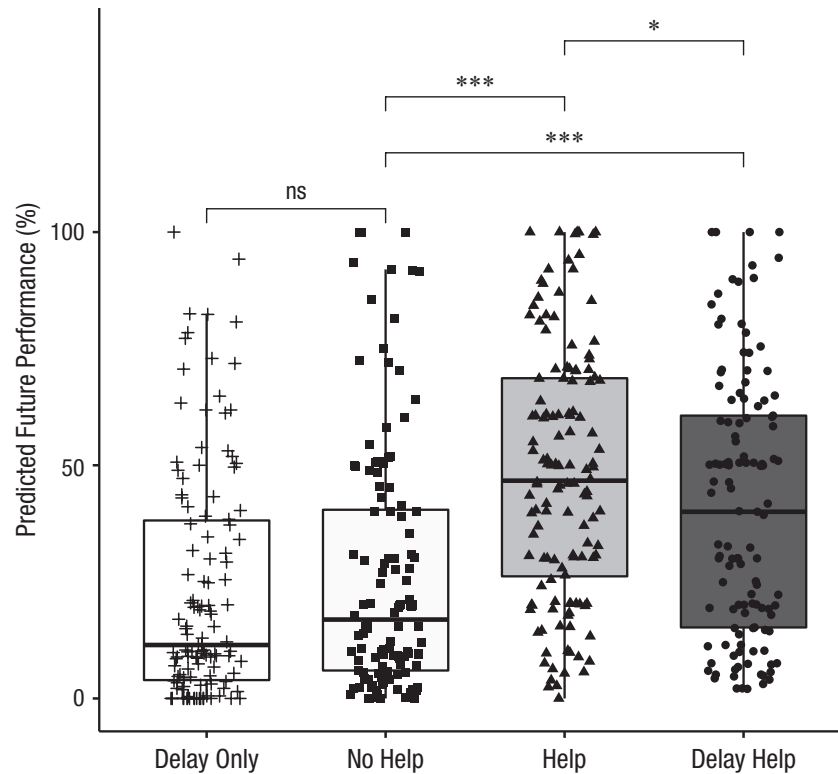


Fig. 3. Predicted future performance (percentage correct) by condition in Experiment 2a. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions ($*p < .05$, $***p < .001$).

As in Experiment 1, participants who had received no help anticipated worse future performance than participants who had received help. Crucially, introducing a delay mitigated but did not eliminate this difference. It is worth noting that the delay neither affected memory nor predictions; the delay-only condition did not significantly differ from the no-help condition. That is, introducing a delay did not lower confidence in and of itself. Instead, it afforded participants an opportunity to query their memories and realize the difficulty of the task. Once again, participants' metacognitive calibration improved when they had experience of what performance would be like without assistance.

Experiment 2b

In Experiment 2b, we again tested episodic memory, except that instead of receiving help after a delay, participants needed to actively click a button to receive help.

Method

Participants. Three hundred one participants from the United States (147 males, 154 females; mean age = 35.79

years, $SD = 11.34$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. The materials and instructions were the same as for Experiment 2a. Each participant was randomly assigned to one of three conditions: no help, help, or button. In the no-help condition, participants attempted to recall as many of the 15 words as possible without any hints. In the help condition, the first three letters of the correct response appeared below each of the text boxes. In the button condition, a button labeled "HINT" appeared above each text box. When clicked, the button would immediately display the first three letters of the correct answer. Last, participants in all conditions were asked, "If you were asked to remember another set of words without any outside help or clues, what percentage would you successfully recall in the correct order?"

Results

Participants in the button condition correctly remembered fewer words from the list ($M = 50.51\%$, $SD = 23.94\%$) compared with those in the help condition ($M = 61.90\%$, $SD = 22.16\%$), $\beta = -0.39$, $SE = 0.12$, $p = .001$, but more

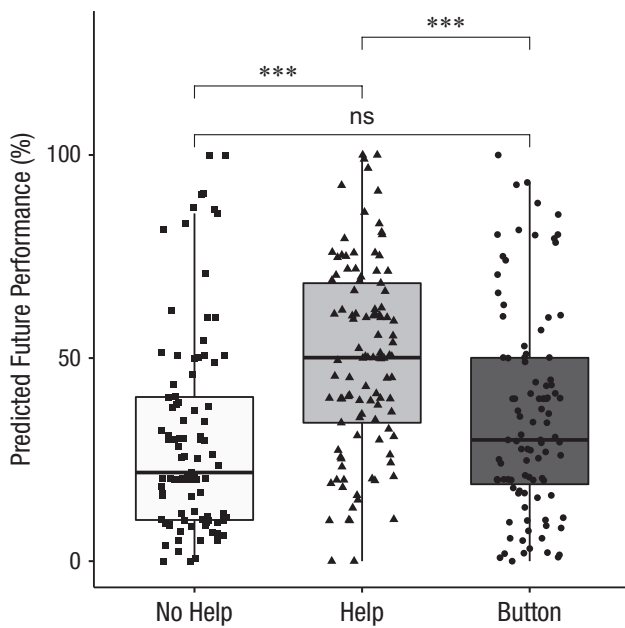


Fig. 4. Predicted future performance (percentage correct) by condition in Experiment 2b. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions ($***p < .001$).

than those in the no-help condition ($M = 24.33%$, $SD = 24.83%$), $\beta = 0.93$, $SE = 0.12$, $p < .001$. Participants in the button condition predicted significantly lower future performance ($M = 35.83%$, $SD = 25.01%$) compared with those in the help condition ($M = 49.94%$, $SD = 23.55%$), $\beta = -0.52$, $SE = 0.13$, $p < .001$, but their predictions were not significantly different than those in the no-help condition ($M = 30.41%$, $SD = 24.75%$), $\beta = 0.23$, $SE = 0.13$, $p = .09$ (see Fig. 4). Sixty-two percent of participants in the button condition clicked to view at least half of the hints. The results remained significant when participants who did not click were excluded from the analysis.

As in Experiment 2a, participants who received help predicted higher future performance than participants who had not received help. When participants had to make an active choice to receive help, in this case by pressing a button, predicted future performance dropped to a level similar to that of those who received no help. People who sought help deliberately, as opposed to by default, experienced the difficulty of the task without assistance (before they asked for the hint) and thus were better informed about how well they would do if the hint were not available.

To this point, we found similar patterns of results for both episodic and semantic memory, but extended cognition has broader scope than just memory. In

Experiment 3, we extended our investigation to problem solving.

Experiment 3a

Although there have been a number of demonstrations that people who off-load memory are subsequently metacognitively miscalibrated (e.g., Fisher et al., 2015), this pattern of results has not been shown in problem solving. Thus, before exploring debiasing interventions, it is important to first establish that a bias exists. In Experiment 3a, we examined people's ability to solve anagrams, with or without hints, and investigated how the presence of hints influences people's metacognitive awareness of their natural abilities.

Method

Participants. One hundred ninety-eight participants from the United States (107 males, 91 females; mean age = 34.73 years, $SD = 10.80$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. To begin, participants were instructed, "In this study, your task will be to unscramble as many words as you can in 90 seconds." Participants next viewed eight anagrams (see Section S2 in the Supplemental Material) while a timer counted down from 1 min 30 s to zero. Each participant was randomly assigned to either the help or the no-help condition. In the help condition, the first three letters for the solution to the anagram were displayed below each item. In the no-help condition, the hints were not displayed to participants. After completing the anagrams (or after time expired), all participants responded to the following question: "If you were to unscramble another set of words without any outside help or clues, what percentage would you successfully unscramble?"

Results

We first examined participant performance on the anagrams. Unsurprisingly, participants who received the first three letters of the solution for each anagram solved more correctly ($M = 65.21%$, $SD = 27.76%$) than those who did not ($M = 21.60%$, $SD = 20.48%$), $t(196) = 12.42$, $p < .001$, Cohen's $d = 1.77$, 95% confidence interval (CI) = [1.44, 2.10]. Critically, participants who had received the hints predicted that they would perform better in the future without hints ($M = 46.24%$, $SD = 26.58%$) than those who did not see the hints ($M = 24.56%$, $SD = 24.65%$), $t(196) = 5.92$, $p < .001$, Cohen's $d = 0.84$, 95% CI = [0.55, 1.34]. The effect of condition remained significant after we controlled for demographic variables (gender, age, education), $\beta = -0.83$, $SE = 0.13$, $p < .001$.

Experiment 3a extended the previous finding beyond memory. People who received hints realized that the hints were helpful, estimating that they would solve 20% fewer anagrams if the hints were not available. However, they nonetheless underestimated how much they had relied on those hints by 25%. Building on this finding, we next explored whether the interventions (delay and deliberate choice) mitigated this effect outside of memory tasks.

Experiment 3b

In Experiment 3b, we again tested anagram solving, with or without hints, but this time we added a condition testing the delay intervention.

Method

Participants. Two hundred forty-three participants from the United States (119 males, 124 females; mean age = 36.18 years, $SD = 12.15$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. As in Experiment 3a, the main task of Experiment 3b consisted of solving the eight anagrams listed in Section S2 in the Supplemental Material. Unlike in Experiment 3a, the anagrams appeared on screen one at a time. Participants were instructed, “In this study, your task will be to unscramble as many of the 8 words as you can in 2 minutes and 30 seconds. Once an answer is submitted it cannot be changed. To skip an item, click the arrow below the word.” Each participant was randomly assigned to one of three conditions: no help, help, or delay. In the no-help condition, participants did not receive any outside aid in solving the anagrams. In the help condition, the first three letters of the correct answer appeared beneath each anagram. In the delay condition, after a delay of 7 s, the first three letters of the correct answer appeared beneath each anagram. A timer counted down the remaining time before each hint appeared. Participants in the delay condition were told in the initial instructions, “For each word, a hint will appear after 7 seconds.” So performance for participants in the delay condition would not be affected, we increased the time limit from Experiment 3a to 2 min 30 s. After the anagram portion of the experiment, all participants answered the following question: “If you were to unscramble another set of words without any outside help or clues, what percentage would you successfully unscramble?”

Results

Participants in the delay condition answered fewer anagrams correctly ($M = 45.29\%$, $SD = 28.41\%$) than those in

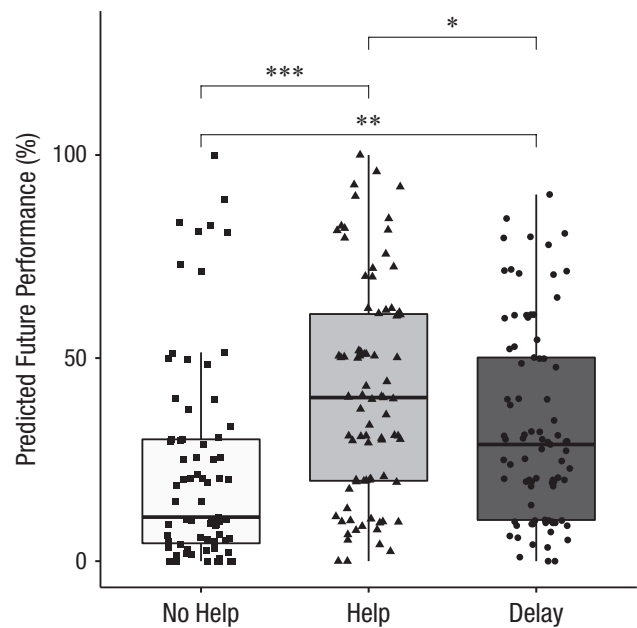


Fig. 5. Predicted future performance (percentage correct) by condition in Experiment 3b. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions ($*p < .05$, $**p < .01$, $***p < .001$).

the help condition ($M = 55.27\%$, $SD = 30.44\%$), $\beta = -0.39$, $SE = 0.14$, $p = .005$, and significantly more than those in the no-help condition ($M = 22.50\%$, $SD = 18.95\%$), $\beta = 0.78$, $SE = 0.14$, $p < .001$. Results were in line with our prediction: Participants in the delay condition provided lower estimates of future performance ($M = 33.19\%$, $SD = 23.95\%$) relative to participants in the help condition ($M = 41.60\%$, $SD = 26.64\%$), $\beta = -0.34$, $SE = 0.15$, $p = .02$, although not at the same level as those of participants in the no-help condition ($M = 22.77\%$, $SD = 25.35\%$), $\beta = 0.40$, $SE = 0.15$, $p = .009$ (see Fig. 5).

As in Experiment 3a, participants who received hints rated their ability to solve anagrams higher relative to participants who did not have help. However, this tendency was partially mitigated in the delay condition. This is consistent with the theory that while people are waiting for hints, they experience how difficult the task is without the hints and thus have better information informing their metacognition.

However, these results should be interpreted with caution, as participants' actual performance differed between the delay and help conditions. That is, in both the help and delay conditions, participants' estimates of future performance (without hints) were approximately 15% lower than their actual performance (with hints). Thus, it could be that the delay does not provide any additional metacognitive calibration above and

beyond the fact that people actually performed worse in the delay condition. This differs from other experiments in which the interventions did not affect performance on the task but rather affected only predictions of performance on future tasks. Thus, although the present findings are consistent with our predictions, they are more difficult to interpret.

Experiment 3c

In Experiment 3c, we tested the second intervention for improving metacognition of the outsourced mind: forcing participants to actively choose to receive assistance.

Method

Participants. Three hundred five participants from the United States (163 males, 142 females; mean age = 35.28 years, $SD = 10.99$) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. As in Experiment 3a, participants were given 90 s to complete eight anagrams. The help and no-help conditions were identical to those of Experiment 3a. Experiment 3c introduced a new condition: the button condition. In this condition, a button labeled “HINT” appeared below each of the anagrams. As soon as participants clicked the button, a hint would appear—the first three letters of the solution, the same hint that participants in the help condition received automatically. As in Experiment 3a, after the anagram portion of the experiment, all participants estimated how well they would perform without any outside help or hints.

Results

Participants in the button condition answered fewer anagrams correctly ($M = 45.54\%$, $SD = 26.31\%$) than those in the help condition ($M = 61.42\%$, $SD = 27.44\%$), $\beta = -0.52$, $SE = 0.12$, $p < .001$, and more than those in the no-help condition ($M = 24.25\%$, $SD = 25.30\%$), $\beta = 0.68$, $SE = 0.12$, $p < .001$. In line with our hypothesis, participants in the button condition gave lower estimates of future performance ($M = 27.44\%$, $SD = 23.98\%$) than participants in the help condition ($M = 48.07\%$, $SD = 27.21\%$), $\beta = -0.69$, $SE = 0.13$, $p < .001$, and estimates in the button condition were no different from those of participants in the no-help condition ($M = 25.07\%$, $SD = 27.44\%$), $\beta = 0.08$, $SE = 0.13$, $p = .55$ (see Fig. 6).

As in Experiments 3a and 3b, participants who received hints provided higher ratings of their ability to solve anagrams relative to participants who did not

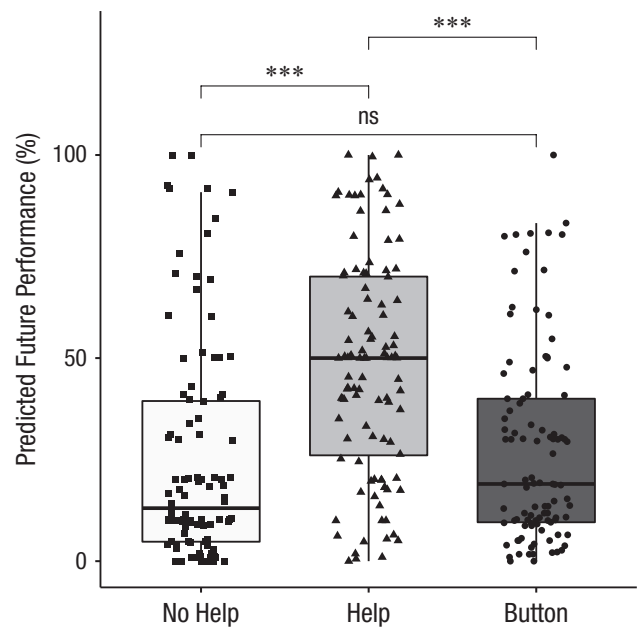


Fig. 6. Predicted future performance (percentage correct) by condition in Experiment 3c. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions ($***p < .001$).

receive hints. Crucially, when they were forced to actively choose to receive the hint instead of having the hint provided by default, this increased self-assessment was entirely eliminated. Unlike in Experiment 3b, this improved calibration went above and beyond the difference in actual accuracy, allowing for straightforward interpretation. Moreover, unlike in many of the previous experiments, in which the bias was mitigated but not eliminated entirely, in Experiment 3c, participants' results in the button condition were statistically indistinguishable from participants' results in the no-help condition. The pattern of results provides strong evidence that the findings extend beyond memory to externally aided cognition more generally.

Experiment 3d

Experiment 3d replicated the results of Experiment 3c and addressed two key questions. First, although participants in the help condition provided higher self-assessments in the previous experiments, we cannot definitely determine whether they were overconfident or whether participants in the other conditions were underconfident. To address this, we assessed participants' accuracy by asking them to complete the task for which they had made predictions. Second, although we have argued that the observed effects are due to

differences in perceived difficulty across conditions, to this point, perceived difficulty had not been directly measured. It is therefore possible that the effects were due to other mechanisms, such as reduced opportunities to learn how to do the task in the absence of hints. Therefore, we measured the perceived difficulty of the task in Experiment 3d.

Method

Participants. Three hundred participants from the United States (165 males, 135 females; mean age = 39.06 years, *SD* = 12.68) completed the experiment online through Amazon Mechanical Turk via TurkPrime.

Materials, design, and procedure. Experiment 3d followed the procedure of Experiment 3c, with several modifications. First, after completing the initial task of solving the eight anagrams, participants were asked, “How difficult did you find unscrambling the previous set of words?” They replied on a Likert scale ranging from 1 (*not at all*) to 7 (*very*). Second, we administered a second set of anagrams in order to assess metacognitive accuracy. We pretested eight new anagrams and then mixed them with the original eight anagrams to create two equally difficult sets (see Section S3 in the Supplemental Material). Each participant was randomly assigned to the no-help, help, or button condition and completed the first set of anagrams just as in Experiment 3c. After predicting their performance for a new set of words using the same measure as in Experiment 3c, all participants completed the second set of anagrams without any help or hints. The order of two sets of anagrams was counter-balanced across participants.

Results

To assess the accuracy of participants’ predicted performance, we calculated overconfidence by subtracting participants’ actual performance on the second set of anagrams from their predicted performance (see Table 1 for predicted and actual performance). These difference scores indicated that participants in the button condition were well calibrated in predicting their future performance ($M = 1.66\%$, $SD = 26.22\%$) and not different from participants in the no-help condition ($M = -1.79\%$, $SD = 32.83\%$), $\beta = -0.09$, $SE = 0.14$, $p = .51$; however, participants in the help condition were significantly more overconfident ($M = 16.08\%$, $SD = 32.19\%$), $\beta = 0.48$, $SE = 0.14$, $p < .001$ (see Fig. 7). These results also align with the pattern observed in the previous experiments. An exploratory analysis examining participants’ predicted future performance showed that

Table 1. Mean Predicted and Actual Anagram Performance (Percentage Correct) in Study 3d

Condition	Predicted performance	Actual performance
No help	30.33% (23.60%)	32.13% (30.79%)
Help	49.75% (27.89%)	33.66% (24.03%)
Button	34.11% (23.13%)	32.45% (22.30%)

Note: Standard deviations are given in parentheses.

participants in the button condition ($M = 34.11\%$, $SD = 23.13\%$) provided lower estimates than participants in the help condition ($M = 49.75\%$, $SD = 27.89\%$), $\beta = -0.62$, $SE = 0.13$, $p < .001$, but predictions in the button condition were no different than those of participants in the no-help condition ($M = 30.33\%$, $SD = 23.60\%$), $\beta = 0.10$, $SE = 0.13$, $p = .47$. Participants in the button condition rated the initial set of anagrams as easier ($M = 5.14$, $SD = 1.39$) than participants in the no-help condition did ($M = 5.91$, $SD = 1.23$), $\beta = 0.43$, $SE = 0.13$, $p < .001$, and, as predicted, more difficult than participants in the help condition did ($M = 4.09$, $SD = 1.77$), $\beta = -0.68$, $SE = 0.13$, $p < .001$. Furthermore, an exploratory analysis showed that the difficulty ratings fully mediated the difference between overconfidence levels of the button and help conditions (bootstrapped

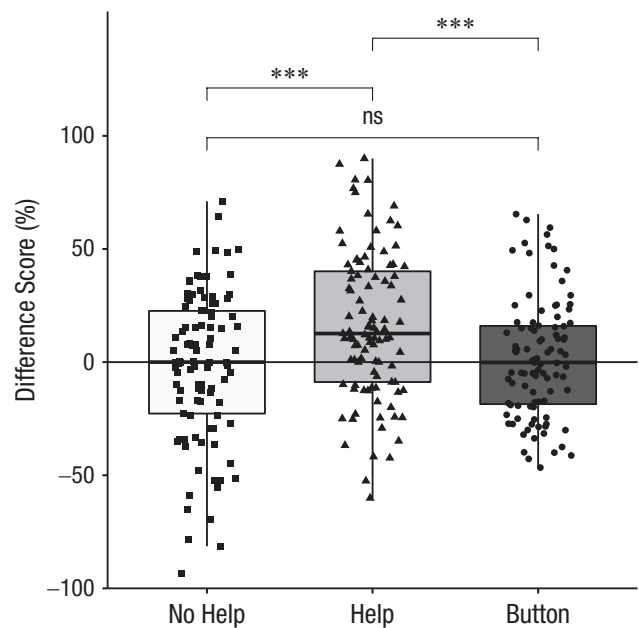


Fig. 7. Difference score (predicted performance – actual performance) by condition in Experiment 3d. The horizontal lines indicate the medians, the top and bottom edges of the boxes mark the interquartile range, the whiskers extend 1.5 times the interquartile range, and the symbols represent individual data. Asterisks indicate significant differences between conditions (***) $p < .001$.

standardized indirect effect = -0.22 , 95% CI = $[-0.34, -0.11]$, $p < .001$; Tingley et al., 2014). Although these results demonstrate the important role of perceived task difficulty, the fact that difficulty ratings but not metacognitive accuracy differed between the button and no-help conditions suggests that other mechanisms may be involved as well. Seventy-one percent of participants in the button condition clicked the button to receive the hint for at least half of the items. The results remained significant when participants who clicked to see less than half the hints were excluded from the analysis.

The results of Experiment 3d replicated those of the earlier experiments and extended the findings in two important ways. First, by assessing the accuracy of participants' predictions, we observed that participants in the help condition were overconfident and those in the other conditions were well calibrated. Second, the results suggest that, as hypothesized, the button increased perceived difficulty, leading to overconfidence in future performance.

Although we did not directly test accuracy in the semantic-memory or episodic-memory task, we found that people's actual memory performance in the no-help condition (which closely mimicked the task for which participants were attempting to predict future performance) closely aligned with their predictions, suggesting that predictions in the no-help condition were more accurately calibrated than those in the help-condition. Furthermore, previous research has shown overestimation of performance when memory is outsourced to assistive devices (e.g., Fisher & Oppenheimer, in press; Hargis & Oppenheimer, 2020).

General Discussion

Across eight experiments, we found evidence for a new theory of how metacognition monitors the outsourced mind and tested two mechanistic predictions of that theory. When assistance is (a) delayed or (b) actively chosen, people more accurately assess their reliance on outside help.

These experiments contribute to an emerging literature on the impact of outsourcing cognition by advancing our understanding in three key ways. First, we showed that very different forms of outside help can lead to similar illusions of knowledge. Previous research examined the impact of powerful technologies such as Google search (Fisher et al., 2015; Ward, 2013), and here we showed that a similar pattern emerges for simple hints (even a single letter). Second, we demonstrated the generality of metacognitive misattribution. Expanding beyond knowledge assessments, we found miscalibration for semantic memory, episodic memory, and problem solving. Third, and most critically, we

identified a cognitive mechanism that makes it especially difficult for people to distinguish their own competence from the outside help they receive. Metacognitive biases become more pronounced when people never experience their own ability without assistance. People use their "feelings as information" (see Schwarz, 2012), and in the absence of feelings of difficulty, they are unaware of how unskillful they are at the task.

The current findings contribute to the understanding of the consequences of distributed cognition. Off-loading can lead to downstream effects, such as forgetting (Kelly & Risko, 2019), reduced effort (Risko et al., 2017), and altered memory strategies (Scarampi & Gilbert, 2020). Our experiments show how external aid can generate overconfidence. Further, we demonstrated how increasing task difficulty through time delays or active choice can help calibrate self-assessments.

These experiments expand recent findings related to how people use new technology such as Google search. Before searching online, people's reliance on Google is shown by their reluctance to provide their own answers (Ferguson et al., 2015). However, after using search engines, people confuse their ability to find information online with what they actually know, leading to increased self-assessments of their abilities (Fisher et al., 2015; Ward, 2013). Furthermore, people judge online information found quickly as more likely to be remembered (Stone & Storm, 2021). Although the current set of findings isolates mechanisms of metacognition within the context of simple tasks, these same principles may also be operating when people use more advanced technologies.

Relatedly, beyond their theoretical contribution, these experiments offer potential practical insights. Increasing use of technology to augment cognition is evident across a wide range of applied environments, including medicine (diagnostic decision aids), architecture (AutoCAD software), engineering (simulation software), education (learning-support technology), and daily life (Fitbits and smartphones). A better understanding of how everyday technologies bias people's assessments of their own abilities can help us predict and prevent errors of overconfidence (Parasuraman & Manzey, 2010). The mechanisms we have identified can help inform the design of future technologies and avoid miscalibration problems.

This set of findings suggests that these effects generalize to other types of outsourced cognition, but cataloging other domains and identifying potential boundary conditions remain promising avenues for future research. Furthermore, these experiments focused on developing a theoretical understanding of outsourced cognition. In future work, these principles and potential interventions could be tested in applied settings.

Transparency

Action Editor: Sachiko Kinoshita

Editor: Patricia J. Bauer

Author Contributions

M. Fisher and D. M. Oppenheimer conceived and designed the study. Testing and data collection were performed by M. Fisher. M. Fisher analyzed and interpreted the data under the supervision of D. M. Oppenheimer. M. Fisher and D. M. Oppenheimer drafted the manuscript. Both authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

Deidentified raw data, materials, and analysis code have been made publicly available via OSF and can be accessed at <https://osf.io/ruxmk>. The design and analysis plans for five of the eight experiments were preregistered at OSF (Study 1b: <https://osf.io/bjyg8>, Study 3a: <https://osf.io/bsw78>, Study 3b: <https://osf.io/5yuv9>, Study 3c: <https://osf.io/ht6gf>, Study 3d: <https://osf.io/x67dv>). This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797620975779>

Notes

1. For all experiments, participants' compensation matched Amazon Mechanical Turk's standard rate of \$6 per hour (based on each experiment's estimated duration).
2. At the end of all experiments, participants reported demographic information.
3. This and all subsequent regression models controlled for the demographic variables of gender, age, and education level.

References

- Alicke, M. D., & Govorun, O. (2005). The better-than-average-effect. In M. D. Alicke, D. A. Dunning, & J. I. Kreuger (Eds.), *The self in social judgment* (pp. 85–106). Psychology Press.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235.
- Alter, A. L., Oppenheimer, D. M., & Zemla, J. C. (2010). Missing the trees for the forest: A construal level account of the illusion of explanatory depth. *Journal of Personality and Social Psychology*, *99*, 436–451.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.
- Block, L. G., & Morwitz, V. G. (1999). Shopping lists as an external memory aid for grocery shopping: Influences on list writing and list fulfillment. *Journal of Consumer Psychology*, *8*, 343–375.
- Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*, Article 45.
- Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2), 149–154.
- Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. <https://doi.org/10.1177/1948550619875149>
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.
- Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*(5), 1080–1127.
- Dunning, D. (2012). *Self-insight: Roadblocks and detours on the path to knowing thyself*. Psychology Press.
- Ferguson, A. M., McLean, D., & Risko, E. F. (2015). Answers at your fingertips: Access to the Internet influences willingness to answer questions. *Consciousness and Cognition*, *37*, 91–102.
- Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*, 939–946.
- Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, *144*, 674–687.
- Fisher, M., & Keil, F. C. (2014). The illusion of argument justification. *Journal of Experimental Psychology*, *143*, 425–433.
- Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science*, *40*, 1251–1269.
- Fisher, M., & Oppenheimer, D. M. (in press). Who knows what? Knowledge misattribution in the division of cognitive labor. *Journal of Experimental Psychology: Applied*.
- Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 680–693.
- Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*, *149*(3), 501–517. <https://doi.org/10.1037/xge0000652>

- Hargis, M. B., & Oppenheimer, D. M. (2020). *You won't know until it's too late: Metacognition and memory offloading* [Manuscript submitted for publication].
- Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, *8*, 31–38.
- Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*, 396–402.
- Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, *193*, Article 104012. <https://doi.org/10.1016/j.cognition.2019.104012>
- Intons-Peterson, M. J., & Fournier, J. (1986). External and internal memory aids: When and how often do we use them? *Journal of Experimental Psychology: General*, *115*, 267–280.
- Kelly, M. O., & Risko, E. F. (2019). Offloading memory: Serial position effects. *Psychonomic Bulletin & Review*, *26*, 1347–1353.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*, 433–442.
- Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology*, *69*, 781–799.
- Nestojko, J. F., Finley, J. R., & Roediger, H. L., III. (2013). Extending cognition to external agents. *Psychological Inquiry*, *24*, 321–325.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410.
- Permut, S., Fisher, M., & Oppenheimer, D. M. (2019). TaskMaster: A tool for determining when subjects are on task. *Advances in Methods and Practices in Psychological Science*, *2*(2), 188–196. <https://doi.org/10.1177/2515245919838479>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*, 676–688.
- Risko, E. F., Huh, M., McLean, D., & Ferguson, A. M. (2017). On the prospect of knowing: Providing solutions can reduce persistence. *Journal of Experimental Psychology: General*, *146*(12), 1677–1693.
- Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.
- Runge, Y., Frings, C., & Tempel, T. (2019). Saving-enhanced performance: Saving items after study boosts performance in subsequent cognitively demanding tasks. *Memory*, *27*(10), 1462–1467.
- Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory*, *28*(5), 677–691.
- Schönpflug, W., & Esser, K. B. (1995). Memory and its “Graeculi”: Metamemory and control in extended memory systems. In C. A. Weaver, III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 245–255). Erlbaum.
- Schwarz, N. (2012). Feelings-as-information theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *The handbook of theories of social psychology* (pp. 289–308). SAGE.
- Stone, S. M., & Storm, B. C. (2021). Search fluency as a misleading measure of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 53–64. <https://doi.org/10.1037/xlm0000806>
- Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, *26*(2), 182–188. <https://doi.org/10.1177/0956797614559285>
- Thomson, K., & Oppenheimer, D. M. (2020). *The (effort) elephant in the room: What is effort, anyways?* [Manuscript under revision].
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5). <https://doi.org/10.18637/jss.v059.i05>
- Tylor, E. B. (1870). *Researches into the early history of mankind and the development of civilization* (2nd ed.). John Murray.
- Ward, A. F. (2013). *One with the cloud: Why people mistake the Internet's knowledge for their own* [Unpublished doctoral dissertation]. Department of Psychology, Harvard University.
- Weis, P. P., & Wiese, E. (2019). Using tools to help us think: Actual but also believed reliability modulates cognitive offloading. *Human Factors*, *61*(2), 243–254.
- Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253.
- Wilson, T. D. (2009). *Strangers to ourselves: Discovering the adaptive unconscious*. Harvard University Press.
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933.